DOCUMENT RESUME

ED 338 673	TM 017 481

AUTHOR	Smith, Mary Lee; And Others
TITLE	The Role of Testing in Elementary Schools.
INSTITUTION	Center for Research on Evaluation, Standards, and Student Testing, Los Angeles, CA.
SPONS AGENCY	Office of Educational Research and Improvement (ED), Washington, DC.
REPORT NO	CSE-TR-321
PUB DATE	May 91
CONTRACT	G0086-003
NOTE	223p.; Pages 54-56 and 216 may not reproduce well.
PUB TYPE	Reports - Research/Technical (143)
EDRS PRICE	MF01/PC09 Plus Postage.
DESCRIPTORS	Case Studies; *Classroom Techniques; Comparative Analysis; Definitions; Educational Improvement; Elementary Education; *Elementary Schools; Elementary School Teachers; Interviews; Psychometrics; School Surveys; *Teacher Attitudes; Test Coaching; Test Results; *Test Use
IDENTIFIERS	Arizona (Phoenix); *External Evaluation; *Testing Effects

ABSTRACT

The goal of this study was to understand the role that external testing plays in elementary schools. Focus was on uncovering teachers' beliefs about testing and preparing students to take tests, how these beliefs and values are organized, and what implications they might have on practice. To accomplish this, the day-to-day life in classrooms and how tests and results come into play were studied. The dual case study design provided an interpretive contrast for two schools from the Phoenix (Arizona) area. Schools used the same external tests (the Iowa Tests of Basic Skills, Basic Skills Test, Continuous Uniform Evaluation System, and Study Skills Test). Although the schools had many similarities, including that of population, one had a program-centered, phonics-based curricular context, and the other had a student-centered, literature-based approach. Observations of 29 classrooms, interviews with 19 teachers, and more extensive observations of 6 focal classrooms made the analysis of beliefs about testing possible and allowed the description of activities related to testing at the two schools, including test preparation and coaching. Study findings are grouped into: (1) local definitions of testing; (2) the role of testing; and (3) the effects of testing. It is held that to define the role of testing as simply psychometric is to oversimplify it, but it is the psychometric weaknesses of tests that make them useful weapons in skirmishes among interest groups. It is argued that no test score ever improves schools. The changes brought about because of test scores are short-term and largely symbolic. Seven exhibits, one figure, and one table are provided. A 70-item list of references is included. Two appendices summarize a survey of Arizona educators and discuss disappointing test scores. (SLD)

C·R·E·S·S·T

Center for Research
on Evaluation, Standards,
and Student Testing

ED338673

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION
CENTER (ERIC)
☑ This document has been reproduced as received from the person or organization originating it
☐ Minor changes have been made to improve reproduction quality
● Points of view or opinions stated in this document do not necessarily represent official OERI position or policy

---

THE ROLE OF TESTING IN ELEMENTARY SCHOOLS

CSE Technical Report 321

Mary Lee Smith
with
Carole Edelsky, Kelly Draper, Claire Rottenberg,
and Meredith Cherland

Arizona State University

UCLA Center for Research on Evaluation,
Standards, and Student Testing

TM017481

BEST COPY AVAILABLE

THE ROLE OF TESTING IN ELEMENTARY SCHOOLS

CSE Technical Report 321

Mary Lee Smith
with
Carole Edelsky, Kelly Draper, Claire Rottenberg,
and Meredith Cherland

Arizona State University

UCLA Center for Research on Evaluation,
Standards, and Student Testing

May 1991

The research reported herein was conducted with partial support from the U.S. Department of Education, Office of Educational Research and Improvement, pursuant to Grant No. G0086-003. However, the opinions expressed do not necessarily reflect the position or policy of this agency and no official endorsement by this agency should be inferred.

Please address inquiries to: CSE Dissemination Office, UCLA Graduate School of Education, 405 Hilgard Avenue, Los Angeles, California, 90024-1521

Table of Contents

Page

Chapter One: Constructing the Research Problem 1

Chapter Two: Beliefs about Testing 17

Chapter Three: Natural History of the Testing Event 65

Chapter Four: Assertions about Testing 184

References 206

Appendices 211

Acknowledgements

We have learned so much from our participants that thanks seem insufficient. It takes special people to withstand the level of scrutiny involved in qualitative research. We are particularly in debt of those five who read and reacted to preliminary reports. They sharpened our understandings in many ways. Thanks also are due to Sandra Mathison, Ernie House, Lorrie Shepard, and our graduate students who read drafts or sections of the report and made helpful suggestions. Gene Glass not only read for substance and ideas but did his best to help with style. For what he could not redeem, the responsibility is the first author's.

Chapter One: Constructing the Research Problem

Although schools have administered standardized tests of achievement for decades, only recently have such tests been used as instruments of social policy. Originally such tests were used to gauge the progress of pupils and compare their accomplishments with nationally representative samples. Tests also gave information about pupil achievement in relation to defined objectives. Teachers could use information from the tests to plan or modify instruction and correct deficiencies. Testing programs, including decisions about what characteristics to measure, which standardized tests to purchase, and how to use the test results were voluntary and controlled by local school districts. Testing was internal.

The decade of the 1970s brought the age of accountability, followed by the decade of school reform. Society's problem was to make public schools accountable for academic outcomes, restructure and change them for the better. By improving the basic skills of school children, reforms would guarantee the nation's economic health. Although many changes have been suggested (longer school days, higher standards for promotion and graduation), the standardized test of achievement has come to be regarded as the most cost-effective means for accomplishing accountability and reform. Increasingly, testing programs became external, that is, mandated by legislatures or governing bodies outside the school districts. The decisions about what knowledge and skills to examine, which instruments to select or construct, and how to use and publicize test results were made centrally rather than locally and were the same for all schools in a jurisdiction. Rather than using the results to rectify problems of teaching and learning, the test results were used to publicize success and diagnose failure of the educational system—school districts, schools, or teachers—and to trigger certain actions and decisions such as certification and promotion.

The argument that test results can improve and reform education rests on a simple, common sense assumption, stated most clearly, perhaps, by James Popham in his article, "The Merits of Measurement-Driven Instruction" (Popham, 1987). He argued that teachers are aware of what external achievement tests cover and focus their efforts on such material. Through the teachers' renewed concentration on this material, usually the basic skills of literacy and numeracy, the pupils learn more effectively. What the tests cover reflects what society values, thus the external testing program has direct and worthwhile effects on the educational enterprise. The external test is a powerful means to a valued end.

Countering Popham's assumption are those who argue either that external tests have few direct effects (e.g., Kellaghan, Madaus, & Airasian, 1982), or that external tests have unintended but injurious effects on what happens in the classroom (Bracey, 1987). For example, valued but untested academic content may be ignored in favor of what the test covers, sacrificing long-term attainment for short-term gains on achievement tests. Or, coaching for the external test may alter the meaning of test results. Cheating may occur if teachers or principals seek to avoid low scores. Whether such effects of testing occur and what to make of them are questions we pursue in this study.

To relate our study to others on the same topic, we highlight three distinctions that recur in the professional literature on achievement testing. First, norm-referenced tests are standardized achievement tests that compare one pupil's performance with that of similarly situated pupils nationwide. Criterion-referenced tests compare a pupil's performance against a defined standard of competence. Although testing professionals often draw sharp distinctions between them and argue over their relative merits, we treat them both as standardized tests of achievement. Both types of tests use closed-ended formats such as multiple-choice items and formal, standard rules for determining the meaning of the responses. Both can be internal or external; both may be used for individual or group assessment.

Another distinction in the literature that serves as background for the review of empirical research is that between high stakes and low stakes testing programs. Madaus (1987) defined a high stakes testing program as one that pupils, teachers, or administrators perceive as likely to have consequences such as grade promotion, graduation, or merit pay for teachers. A low stakes testing program is one without such perceived consequences, as when a state merely provides test data to districts so that they can diagnose or fix their own problems as they see fit.

Scholars of applied testing have distinguished several varieties of test preparation. Mehrens and Kaminski (1988, pp. 10-11), for example, listed the following, of which at least the last four they deemed unethical:

1. General instruction on objectives not determined by looking at the objectives measured on standardized tests;
2. Teaching test-taking skills;
3. Instruction on objectives generated by a commercial organization where the objectives may have been determined by looking at objectives measured by a variety of standardized tests;
4. Instruction based on objectives (skills, subskills) that specifically match those on the standardized test to be administered;
5. Instruction on specifically matched objectives (skills, subskills) where the practice (instruction) follows the same format as the test questions;
6. Practice (instruction) on a published parallel form of the same test;
7. Practice (instruction) on the same test.

Our stance has been to use these distinctions and issues as working hypotheses and background knowledge and let the ideas that hold sway in our research site come to the fore. Concerning the question of the ethics of training pupils for taking tests, we take a value-critical position (Rein, 1976). We wish to bring to the surface those beliefs about testing and preparing pupils to take tests that exist in the minds of educators and show how these beliefs and values are organized, and what implications holding them might have on practice.

The Research Literature on the Effects of Testing

The research literature on the effects of external testing is small but growing. In this section, we present the findings of those studies that bear most closely on our own.

In their study of testing practices, Dorr-Bremme and Herman (1986) found that internal testing and external testing were "functionally independent" (p. 95). Internal tests were useful to teachers in supporting instruction and evaluating pupils, whereas external tests, either norm-referenced or criterion-referenced competence tests were not. Teachers believed that external tests, on the other hand, had changed their instruction so that they focused more arduously than before on the general content of the tests. Sixty-two percent of their sample of elementary teachers reported that tests of minimum competence had increased the time they spent on the content domains covered by the mandated testing. Eighty-eight percent reported that basic skills teaching, including remedial work, consumed a significantly greater proportion of their school's time and energy as a result of the mandated testing. Forty-six percent said they were spending more time preparing pupils to take mandated tests. About the same percentage claimed that pressure from mandated testing has a generally beneficial effect. But three-quarters of them said that districts and states should not hold teachers accountable for their pupils' test scores. Among other things, the authors concluded that, when the stakes are high, external tests influence curriculum to a greater degree than in states and districts where consequences are few.

Teachers interviewed by Edelman (1981) reported conducting specific activities to prepare their third grade pupils to take the state mandated test. Sixty percent emphasized the content of the tests over a significant amount of time, while 30 percent covered items specific to the test. Twenty percent of the teachers said they were pressured to change what and how they taught, and almost 40 percent believed that they would be evaluated based on the test scores their pupils received.

Focusing their attention on the role of mandated testing in mathematics, Romberg and his colleagues (Romberg, Zarinnia, & Williams, 1989) surveyed teachers and found that more than 80 percent had changed their teaching in response to the test of math achievement (either norm- or criterion-referenced) that their districts and states mandated. As a result of the external test, teachers concentrated their efforts on basic skills and paper-and-pencil computation. Teachers said that they tailored the topics they cover to fit the content of the test. As a result of the test, they had changed their teaching methods as well, toward direct instruction rather than discovery. They reported that their attention to instruction in problem-solving had increased, but their notion of problem-solving was solving story problems and teaching tricks that students can use to convert story problems to computations. What had diminished because of the test, according to the teachers, were project work, activities involving calculators and computers, topics not emphasized by the test, and cooperative learning activities. In states and districts where stakes were high, teachers were more apt to modify their content and teaching methods to conform to the test. The authors concluded that the impacts of external testing contravene the recommendations of the National Council of Teachers of Mathematics, which advocates cooperative learning, project work, use of computers and calculators, and problem-solving of a creative sort.

In their interviews with teachers, Darling-Hammond and Wise (1985) uncovered a dissonance between the means and the ends of policies to reform education by mandating standardized tests. The majority of those interviewed believed that recent increased emphasis on test results has changed their teaching. Some regard this as ameliorative. For them, the test provides a target and a set of expectations about what various constituencies consider important content and significant attainment. For most other teachers, however, the imposition of standardized testing causes them to slight previously valued areas of the curriculum and real-life skills that tests fail to cover. To the teachers in this study, the same effects hold for criterion-referenced tests as for norm-referenced standardized achievement tests. Both sets of tests attempt to measure in a common way what is intrinsically variable and uncontrollable: the different traits of pupils and different resources available to the teachers. As Darling-Hammond and Wise interpreted it, the means (test scores) become substituted for the ends (genuine attainment). Nor are the ends agreed upon by consensus among the policy-makers who mandate the tests and the teachers who must be responsible for attaining the standard laid down. The authors concluded that society should monitor the intended and unintended effects of mandated testing.

Comparing two states, one with high stakes and one with low stakes testing programs, Wilson and Corbett (1989) concluded that there was a relationship between the perceived effects and the perceived power of the programs. In the state with high stakes testing, educators believed that schools had become focused on an all-out effort to improve test scores and were less concerned with building their general capacity to improve education. Compared with the low stakes state, the high stakes state made more adjustments to its curriculum and instructional programs, changes which educators there believed had narrowed the curriculum and improved it. After the test was imposed, they perceived a greater disparity than before between what the schools taught and what the teachers valued. The authors concluded that "...at some point during an increase in stakes, a shift in local focus occurs, and student performance becomes an end in itself rather than merely an indicator of student attainment of broader learning outcomes" (Corbett & Wilson, 1989, p. 1).

In her case studies of testing effects, Mathison (1987) delineated the types of effects external tests had. New curricula (texts or supplementary materials) were adopted in an attempt to raise low scores. Skills and concepts that were included in standardized tests were added to the curriculum or, if already present, were given greater weight, and those not on the test were de-emphasized. In extreme cases, the test itself became the curriculum for a course. Forms of instruction mimicked forms of the test (e.g., instruction in spelling became a matter of the pupils looking for spelling errors in standardized lists of words, a format employed by the standardized test). Teachers spent more time reviewing material previously covered, more intensively and frequently as the test approaches. Teachers started to group pupils more often by ability, in the belief-sumption that ability grouping produces higher test scores. The sequence of content was altered to conform to the schedule of standardized testing. Remedial courses were added to the curriculum. Formal courses or units on techniques of taking tests were added to the curriculum. When test scores were used to place students in different tracks, a differentiated curriculum was created. Mathison also concluded that these effects vary along six interrelated dimensions: degree of administrative interest, reporting format (by pupil or group), type of test used, grade level, subject matter, and relative power of testing to produce consequences such as teacher merit pay or pupil promotion.

In her analysis of the effects of educational reforms using competency testing programs, Ellwein (1987) found that such effects turned out to be more symbolic than real. Although policy-makers intend that reforms such as promotional gates testing or setting cut-off scores on tests for graduation will raise the overall level of educational attainment, conditions such as the ease of the test or cutoff scores, the erecting of safety nets such as remedial programs for those who fail, and the allowance of multiple retakes of the originally failed tests mitigate the intended effects of the reform. An obvious effect of the mandated test on the elementary school case was the creation of remedial programs. For children who fell below the cutoff for passing from kindergarten to first, the school developed a skill-oriented remedial program so they might pass the test the next time.

According to Airasian and Madaus (1983), in high stakes environments teachers will coach their pupils on test-taking strategies, on the content domain covered by the mandated test (assuming the teachers are aware of the content), and, where security is lax, on the test items themselves. Many empirical studies bear on the success of coaching. The synthesis of research by Kulik, Kulik, and Bangert (1984) showed that taking practice forms of the test can increase test scores. More frequent testing produces greater gains on the final criterion measure (Bangert-Drowns, Kulik, & Kulik, 1988). Meta-analyses of training in test-wiseness and other means of preparing pupils for tests (Samson, 1985; Scruggs, White, & Bennion, 1986) confirm that such training increases test scores. Studies of commercial programs designed to raise scores (Deaton, Halpin, & Alford, 1987) have been negative or inconclusive. According to Mehrens and Kaminski (1988), however, use of programs that tie their activities to specific achievement tests is roughly equivalent to practicing on actual items from the current or recent form of the test itself.

Relationship of the current study to previous research. Past studies of the role and effects of testing contribute working hypotheses for this one. The basis for the evidence so far produced on the topic of effects of testing has been the interview and survey of the beliefs about testing held by various constituencies within the educational system. Past researchers have not examined the classroom directly for traces of testing effects. In the next section we describe the conceptual framework and methods of the present study.

The Study

Conceptual Framework

The goal of the study reported here was to understand the role that external testing plays in elementary schools. To accomplish this end, it was necessary to examine intensively and extensively the day-to-day life of classrooms and schools and to study how tests and test results come into play. Qualitative research is the methodology of choice in such a pursuit. Best outlined by Erickson (1986), qualitative methodology requires that a researcher (a) establish long-term relationships with the people (hereafter called the participants) they study, (b) collect extensive amounts of data by various methods over a long period of time,

(c) observe directly the actions of the participants, (d) understand the participants' meanings by observing and interviewing them, (e) conduct a rigorous analysis of the data gathered, and (f) render the results of the study in such a way that readers have a vicarious experience of the participants, setting, and the phenomena of interest. Qualitative research is based on the assumption that social action cannot be understood apart from the context in which it occurs, and that the context includes the meanings and intentions of the participants, the historical sequences in which actions take place, and the organizational and cultural milieu.

Symbolic interaction (Stryker, 1980) is the theory of social life that informs this methodology; the individual's definitions of the situation and the immediate social context are continually interacting and shaping one another. One cannot know individual meaning perspectives about the role of tests without referring to the immediate social context from which such perspectives arise. From the symbolic interactionist conceptual framework, the researcher aims to understand how individuals' definitions of the situation, worked out in social interaction, contribute to their actions, that is, to their behavior and the intentions underlying their behavior. In addition, the researcher studies the institutional rules and norms that educators may take into account in their actions concerning testing.

To understand social action in context requires that the researcher examine a few settings intensively and forsake the kind of survey design that might yield statistically generalizable but superficial findings. Qualitative research employs a wide variety of methods and controls observer effects, reactivity, and mono-method bias. The validity of the report rests on (a) the adequacy of the relationships developed between researcher and participants, (b) the extent and adequacy of the data produced, (c) the variety of methods used to gather data, (d) the adequacy of the analysis in terms of identifying recurrent patterns of action and meaning, (e) the rigorous search for evidence that disconfirms those patterns, and (f) the credibility of the eventual accounts to readers and participants. The reader plays a role in establishing the validity of the study. If the descriptions and assertions convince the reader of the rigor of the data collection and analysis and the verisimilitude of the account, if they illuminate the phenomena of testing in the schools and change the reader's way of thinking about the issue, then the account is valid.

In summary, the goal of the study is to understand the role of testing in elementary schools by intensively examining testing activities in two schools. More specifically, we sought to document the range of testing activities and understand how these activities are organized into testing events (e.g., all the activities that make up planning for, taking, and reacting to the testing events themselves and results of the test); relate external to internal testing; understand the impact of external tests on what is taught, the methods by which it is taught, and the organization of the schools; understand the meaning of tests to teachers and others, and how these meanings are organized; and describe and interpret the phenomenon of test preparation.

Methods of the Study

Sites selected for the study. The study employed the dual case study design to understand the phenomena of testing in elementary schools. Each school provides an interpretive contrast for the other. Two schools from the same district (Cactus District is the pseudonym we use) were chosen from many in the Phoenix metropolitan area to participate in the study. By choosing two schools from the same district, we made sure that both schools experience the same institutional structures and demands. The two schools share common external tests: the Iowa Test of Basic Skills (ITBS) is a state-mandated test given in April to all pupils in every grade each year. The district-mandated tests include the Basic Skills Test (BST), which is given in May to all pupils in grades three through six, and the Continuous Uniform Evaluation System (CUES), which is given periodically and reported to the district office three times each year. Both the BST and CUES are objectives-based, mastery-level tests that are required by the State of Arizona but developed and administered independently by each school district. The Study Skills Test (SST), mandated and developed by the district, is given in the spring to pupils in grades three through six. The district publishes a scope and sequence of topics and list of required textbooks that district administrators expect the schools to follow. Both norm-referenced tests such as the ITBS and criterion-referenced tests (CUES and BST) are defined as standardized in the sense that the same test items and answer options are given to pupils within a defined testing population, and the results are interpreted by using prescribed statistical procedures. For the most part, the formats of standardized tests provide questions and multiple options from which the pupils must recognize the single correct answer.

The environment of the two schools is high stakes (Popham, 1987) or high power (Mathison, 1987). According to Mathison, a high power testing program is one in which the results of testing are "used for purposes that would have significant consequences: for example, curriculum evaluation, grade-to-grade promotion, teacher or principal evaluation, and funding allocation" (p. 39). Cactus in particular and the State of Arizona in general meet several of these criteria. In keeping with the conceptual and methodological frame of reference adopted in this study, we consider the meaning of high as opposed to low stakes or power from the perspectives of the participants.

Although Cactus School District is relatively centralized in curriculum, standards, and operations (according to our experience with districts), both the schools involved in this study have, nevertheless, what amounts to variances from the district to follow unique curricula in reading. Jackson Elementary School follows a Whole Language approach in many of its classrooms; Hamilton Elementary School (both pseudonyms) follows Reading Mastery (a Direct Instruction program) throughout. The choice of these two schools permits the interpretation of the role of external tests in different curricular contexts: program-centered, phonics-based on the one hand, and student-centered, literature-based on the other. Both schools are alike, however, in that they serve mixed ethnic and predominantly low income populations. On the list published by the State Department of Public Instruction of the ITBS average grade equivalent scores of every school at every grade, both schools are below the mean. In a district with substantial variability in measured achievement from school to school, Jackson and Hamilton are ranked near the bottom.

The schools are also alike in that they have outstanding leadership. Both principals are young, bright, well-educated, caring, and hard-working. Both are deeply committed to their own philosophies of schooling, these philosophies being at variance with one another.

Access to schools. Jackson was chosen for this study because two of us had previous contact with Mrs. Mitchell, its new principal. We selected Hamilton because it was part of the same district and had a similar pupil population but differed in its curriculum.

Access to the schools for the project was sought at school and district offices, proposals were submitted and discussed, and permission to do the study was obtained from all the principal parties. The participation of teachers within the schools was sought through letters and oral presentations at staff meetings early in the school year. At each stage in the negotiations, the project plan was described and explained, questions answered, and confidentiality promised. Teachers were told that their participation was to be voluntary and were promised project reports.

Throughout the project, relationships between researchers and staff were smooth and cordial. No request for information or time was refused. Some teachers at Hamilton suspected that our actual intent was to conduct an experiment pitting Whole Language against Reading Mastery, with ITBS scores as the criterion, but we did our best to disabuse them of this idea. Although our practice is more consistent with the Whole Language philosophy, our research interests for this study were on testing and its role in school life. Curriculum was part of the context of the study in the two schools, not our primary focus. Despite their reservations, the teachers were willing to accept us based on our promises and give us the benefit of the doubt. As the year went on, we became part of the social scenes at the two schools, and close and trusting relationships developed with several teachers. At the end of the second year, we presented preliminary reports to the focal teachers. Two of the four teachers requested that we later alter some of the descriptions of their classrooms. We were able to accommodate requested changes without violating the portrayal of the role of testing in these schools.

To compensate the staff for taking part in the study, we looked for specific things they needed. Hamilton staff expressed the need for a camera, and one was purchased for them. A used computer was also donated to the intermediate grade teachers for use in their record-keeping. In the year following data collection, we assigned a graduate assistant from our university to Jackson to help with their new discipline program.

Observation of school life. The design of the study called for purposive sampling of classrooms for observation during the fall semester. Our plan was to observe on one occasion at least two classrooms per grade at each school. Three purposes lay behind this plan. First, we wanted to get an overview of school life at a time of minimal external testing. Second, we wanted to have a basis in observations for interpreting the statements of teachers whom we would later interview. Third, we needed to select a subgroup of teachers for interviewing and more intensive observation during the second semester. The authors carried out this design, eventually encompassing the day-long observation of 29 classrooms. Detailed notes were taken by hand and transcribed from audio tapes. Notes were descriptive and interpretive, and aimed at building an understanding of ordinary instruction and curricula in the two schools. In addition, we observed and recorded staff meetings and meetings of Team-Assisted Planning (TAP), where teachers refer pupils for consultation, evaluation, and modifications of program or placement.

Interview methods. Identification of a subset of teachers for interviewing occurred in January. Of the twenty teachers we asked to participate, only one refused.

In designing an interview agenda, we followed several principles. First, we assumed that teachers' knowledge and beliefs can best be characterized as personal or tacit rather than propositional in form (Feiman-Nemser & Floden, 1986). To illustrate, knowledge in propositional form is something like the following assertion: "To be valid, a test must measure the content taught in proportion to the extent it is taught in the content universe." Personal knowledge is more likely stored and reported in the form of stories such as the following: "I remember last year when we were administering the ITBS, and one little girl got so upset that she got a pencil stuck in her earring. And it was because there were questions on there about multiplying, and we had never got further than subtraction with regrouping; and I thought it was just so unfair." Second, we assumed that such personal knowledge is best ascertained through soliciting examples and stories from teachers than inferring knowledge and beliefs from this case knowledge (Smith & Shepard, 1988). Clinical interviewing methods are best suited to these principles about the nature of teacher beliefs and knowledge and the ways to elicit them. In clinical interviewing, the researcher starts with an agenda, or list of general topics to cover, as well as an opening statement and open-ended question designed to elicit the participants' perspectives without sensitizing the participants to any hypotheses of the researcher. The content, feeling, and word choice of the participants' initial response then become the structuring mechanisms for the next phase in the interview. As the interview progresses through mutual negotiation, the researcher's agenda is covered naturally. If not, in the latter stages of the interview, more direct questioning can broach the remaining topics. Methodologists of interviewing and narrative psychology such as McCracken (1988), Mishler (1986), Polkinghorne (1988), and Sullivan (1954) have been helpful in developing our ideas about interviewing methodology.

For this study, the authors developed the interview agenda after three months of observing classes and meetings in the two schools and talking informally with teachers and administrators throughout that period. The resulting interview agenda covered four topics: teachers' perceptions of the validity and utility of the external tests that they are required to give, the effects of tests and test scores on teachers, the methods for preparing pupils to take external tests, and the effects of test-taking and test results on pupils.

The tactics of interviewing were the following: An opening statement assured confidentiality and the researchers' neutrality with respect to the topic. The orienting question for the first topic was, "Does the ITBS tell you anything about pupil achievement?" Depending on the initial response, subsequent questions followed the teacher's lead, exploring, for example, what the ITBS does or does not measure, the ways in which that teacher uses test scores, or the perceived reasons that the ITBS is poorly matched with local programs or pupils. Thereafter, the interviews proceeded in nonstandardized ways, depending on the interests and direction of each teacher, but always with an eye to completing the agenda. That is, similar ground would be covered about the BST and CUES, and what information each external test supplies about pupil achievement, teacher and program effectiveness.

The interviewers provided such reassurance as was needed that the interviewers considered the teachers experts in their own classroom and that their perceptions and beliefs were valued. In every case, we took care not to prejudice the responses by asking leading or technical questions such as, "Does the BST have adequate test-retest reliability?" Instead we explored cases and incidents in the teachers' experience.

On the second topic, the orienting question was, "What are some of the things that go through a teacher's mind when she (he) sees the scores on the ITBS from her (his) class?" A question phrased this way does not assume or suggest an assumption to the teacher that test scores have effects, negative or positive. For every initial response, follow-up questions probed for specific incidents, memories, and particulars that would permit inference about the teachers' meanings.

The third topic was opened by asking, "The ITBS is given in April. At what point do you start thinking about them and talking about them to pupils?" If teachers acknowledged preparing for testing, we later asked them to describe what they did to prepare at various stages prior to the testing event. Possible follow-up leads concerned familiarity with item formats, test content, access to previously administered tests, and altering curriculum or methods of instruction to accommodate to tests.

Opening the fourth topic was the question, "In your experience with giving tests such as the ITBS, have there been any effects on pupils that you have noticed?" Because of our informal discussions with teachers, we knew that teachers were particularly sensitive to this issue. Therefore we decided to open this topic last so as not to slight the other topics. Except for this latter condition, the sequence of topics was not uniform, so that we could enhance the opportunity for the participants' interests and language to structure the interviews.

The senior authors of this study (Smith and Edelsky), both highly trained and experienced qualitative researchers and interviewers, conducted the interviews. Averaging one hour in length, the interviews were conducted before or after school or during preparation periods, and always in the teacher's classroom. Without exception, the teachers were cooperative, entered into a collaborative spirit, and were interested in and knowledgeable about the topics. Good rapport was maintained throughout, and several teachers took the opportunity of having a good listener to talk about many issues of concern to them. At the close of the interviews we asked the teachers to sign a consent form that promised confidentiality of site and informant, informed consent, non-coercion, and other requirements for the protection of human subjects.

Observation of focal classrooms. Among those teachers interviewed, we invited six to participate in the extensive observations designed to take place during the spring semester. These included one second, third, and sixth grade teacher at each school. Although initially agreeing, the third grade teachers at both schools, and one sixth grade teacher, later asked to drop out of the study, finding the extent of observation called for in the design to be too intrusive. In the end, this phase of observation involved one second and one sixth grade teacher at each school. These teachers were given a small honorarium for their efforts. Each one chafed a bit at the exposure but persisted until the end of the school year. These teachers were observed for full days either once, twice, or three times per week depending on the week's proximity to the tests. The researchers, principals, and focal teachers negotiated the final observation schedule. In all, we spent a total of 81 days observing these classes.

During the observations of the focal classrooms, the role taken by the researchers was that of "more observer than participant" (Gold, 1958). Although in contact with the participants, the researchers did not act as teachers, evaluators, monitors, or aids in the classrooms. When teachers or pupils initiated conversations or asked for help, the researchers responded as would friendly adults. But teachers' requests for feedback on their instruction were met with polite demurs. Researchers attempted to communicate their attitude that teachers and administrators are experts in their own sphere of activity and that the researchers are not there to reform, criticize, or praise, but only to study. On several occasions, we reminded teachers that we would not report anything we observed in classrooms to administrators or other teachers. We took pains to avoid identification with any one faction or part of the school organization or disturbing the existing authority relations within the school.

During the days spent in the classrooms, researchers recorded as many concrete details as possible of what was taught, the teaching methods by which it was taught, who was in the room or pulled out of it at any time, the time allocation and sequence of events, language and interactions between teachers and pupils and pupils with each other, materials used and unused, intrusions into the classroom by itinerant teachers, parents, aids, and administrators.

To make a durable archive of the ephemeral activities of the everyday life of these classrooms, the researchers made detailed handwritten notes and audio tape recorded as much as possible. Maps and charts of physical spaces, objects within them, and their relationships to actors and events in the classroom were kept. Worksheets, tests, and other materials were collected, textbooks were examined, and relevant portions were photocopied. During their free time, we questioned teachers about the purposes of the classroom activities we observed and their estimates of how typical or unusual events of the day had been. Pupils' reactions to events were recorded. During times when pupils were at recess, lunch, or special classes, researchers either stayed with teachers or went to the teachers' lounge to relax or listen to teachers talk informally about their concerns.

At the end of the school day, researchers transcribed notes and tapes into the more permanent and readable form of write-ups. To these we appended documents and interpretive commentary. Photocopies of the write-ups were provided to the rest of the researchers on the project so that feedback could be given about accuracy, level of detail, and completeness of write-ups as well as reflections on each other's interpretations.

The purposes of these observation activities were to understand everyday classroom life and "ordinary instruction" (which we defined as the contents and methods of teaching relatively unaffected by external tests), document the relationship between external tests and internal tests and methods of assessment, and record the activities of preparing for the external tests and the trends in classroom activities before, during, and after the external tests themselves. The comprehensiveness of the observation, the close access to the everyday life of these classrooms, the detail in recording form the bedrock of credibility of the study itself.

Other methods of data collection. Although we did not request access to the pupils, some teachers solicited comments from them about their perceptions of tests. Two teachers provided short essays or journal entries written by pupils in response to directions to write about the ITBS. In two classes we were able to conduct and record group discussions about pupils' reactions to tests.

One week after the ITBS testing, we conducted a group interview with teachers at Jackson Elementary, using as stimulus alternative forms of the test. The seven teachers participating read and reacted to parts of the test, evaluated it for appropriateness, clarity, and validity and identified specific problems in the content of the tests. This interview was tape recorded and transcribed.

Three of the four focal teachers were interviewed again after the results of the ITBS and other tests became available. We asked them for their reactions to these scores and how the scores might have confirmed or overturned their predictions about the performance of individual pupils. These interviews were recorded and transcribed.

District and building administrators and specialists also made themselves available for formal interviews and many informal discussions. The agenda and tactics of the formal interviews were the same as those used in the interviews of teachers.

While the observations were being made, we collected a rich store of documents: tests, test score reports, district directives about testing and test scores, curricular materials, pupils' products, agenda from meetings, newspaper articles, and district and school newsletters. In addition, we collected articles from the professional and educational trade papers that reflected various ideologies about external testing. For example, we collected newsletters from FAIRTEST, an organization critical of tests, and statements from political advocates and adversaries of external testing.

By the end of the period of data collection, which went on from August, 1987, to October, 1988, we had accumulated nearly 2,000 pages of observation write-ups and documents, in addition to interview transcripts. Multiple copies of the data record were circulated among the observers so that they could add commentary and raise questions and issues.

Analysis of data. Analysis of qualitative data proceeds simultaneously with data collection, with working hypotheses generated at one stage of the study informing the design at subsequent stages. The interpretive portion of the field notes contained suggestions about what the actions and statements of participants relevant to testing might imply. At the end of the period of data collection (October, 1988), formal analysis of the completed data record began, starting with the interviews of teachers.

Interviews. Working with transcriptions of the tape-recorded interviews, the first author followed the procedures of grounded theory methodology (Strauss, 1987). The text files were prepared for analysis by Ethnograph (Seidel, Kjolseth, &

Clark, 1985), a computer program for the qualitative analysis of textual data. Ethnograph allows the analyst to attach codes (the labels of categories) to segments of the text so that the data can be efficiently searched and sorted into the categories or ideas embedded in them.

Working with the text files, the researcher engaged in open coding, the initial activity in grounded theory analysis (Strauss, 1987). Open coding consists of reading the data intensively, line by line, and identifying the ideas and meanings that might be in each line. The researcher makes notes in the margins about these ideas, about what each statement might imply. This process is not simply one of attaching shorthand symbols to topics, but a way of identifying concepts in the data.

Open coding leads to a list of categories for analysis. Categories may refer to topics (e.g., statements about CUES), assumptions and beliefs (e.g., statements about the unreliability of CUES), or social processes and interactions (e.g., excerpts that show how teachers ritualize the administration of CUES).

Interrupting coding to write memos is a key element in grounded theory analysis. A memo represents the analyst's current interpretation of the meaning of a particular category and suggests possible connections with other categories and to the analysis as a whole. An example of a memo is reproduced here. It was written following the coding of the second text file and refers to the beliefs about validity held by a particular teacher.

> MEMO 9/12
> According to _________, a test is valid if it accurately measures what a child knows on a given day. This is a property of the category VT1—Beliefs about that which test scores indicate.
>
> This also pertains to RTTEMP, the category relating to the error of testing due to the temporary characteristics of test-takers.
>
> But to her, achievement itself seems to be a transitory thing. A child may "have it" one day and not the next. Tests only measure the one-day state and not any enduring trait of competence of child, teacher or program. This suggests a new category to be added: The dimensions of beliefs about the nature of achievement—ACHVMNT.
>
> It seems likely that there is a connection here between this category and the psychological and social DISTANCING that goes on among teachers who perceive themselves as trying hard but still having their pupils score low.

Focus coding consists of using a list of categories (generated inductively during open coding) and attaching them to sections of text that contain relevant ideas. During focus coding the analyst is alert to new ideas for which new categories are created. Text files already coded are then combed for instances of the new category. In this study, the category "Asserting alternative and untested goals for schools (ALT-GOAL)" was discovered and named during the focus coding of the second group of transcripts.

Constant comparison analysis involves the coding of each incident (segment of the data from the text file) into as many categories as possible and systematically comparing every incident coded within a category with all other incidents coded in that category. By examining the incidents coded in the same category, the analyst discovers the dimensions and properties of that category and converts that discovery (in the form of memos) into category definitions, properties, conditions, and consequences. One also begins to interpret the category in terms of its relationships with others. In addition, the meanings of categories are clarified, some becoming the properties of other categories.

The analytic sequence proceeded as follows: open coding and memo writing for the first transcript, development of a list of categories, focus coding of transcripts two through four, constant comparison analysis and memoing of the accumulated files, open coding and memoing of transcript number five, focus coding of transcripts six through eight, constant comparative analysis and memoing of one through eight, open coding and memoing of transcript number nine, focus coding of ten through twelve, constant comparison analysis and memoing, focus coding and constant comparison analysis of the remaining transcripts. The order of analysis of transcripts was arbitrary, except that transcripts of teachers from the two schools were thoroughly mixed.

The next stage in the analysis was the search for the structural relationships among categories and properties and the discovery of the core category. As described by Glaser (1978), a core category is the main theme or "concern or problem for the people in the setting" (p. 95). Characteristics of core categories include explanatory power, centrality ("be related to as many other categories and their properties as possible and more than other candidates," p. 95), frequency of occurrence, and variability. Meeting these criteria was obvious for the category, "Defining the discrepancy between the indicator and the trait of achievement." Constant comparison analysis of this category revealed the centrality and explanatory power of this concept relative to much of the data. It occurred most often and occurred in conjunction with more other categories than any other candidate. In Chapter Two, we present this argument in detail and the data that support it.

We cross-referenced the evidence from teachers' interviews with statements they made during our observations of classrooms, staff meetings, and lounge conversations, constituting a kind of multi-method triangulation. We also asked a small group of participants, experts in testing and qualitative research, and students in an advanced course in qualitative research to read and react to drafts and portions of the analysis. In addition, we obtained validation when a survey of Arizona teachers on the same topic produced findings similar to ours. We describe this study (Nolen, Haldadyna, & Haas, 1989) and summarize its results in Appendix A.

The analysis of administrator interviews used focus coding procedures. Following the principles of theoretical sampling (Glaser, 1978), we looked for new properties of the core category discovered in the teacher interviews and new categories that might serve to sharpen our theoretical understanding of the concepts and allow us to hypothesize about the relationships among categories and organizational roles that administrators occupy. We selected segments of data that illustrate these concepts and assertions. However, at this point in the analysis, we felt that the small number of administrators participating might reveal their identities. Therefore, we elected to conduct interviews with additional administrators in the same and other districts, add these data to the other, and select significant excerpts from the larger collection. We took care to make sure that the meaning of the categories was not compromised by this process. Chapter Two contains the excerpts from this analysis.

To study pupils' beliefs about testing, we recorded their comments about tests made during our observations of their classrooms, obtained secondhand information from their teachers, conducted group interviews with a fifth grade class in one focal school and a sixth grade class in the other, and collected journal entries on the subject of testing. We attempted to sample theoretically to generate further properties of the core category and contrast pupils' beliefs with those of teachers. To present significant excerpts, we edited together parts of their interviews and represented them as one. Because of the consistency of their statements, no bias was introduced by this procedure. Of the journal entries of the primary pupils, content analysis revealed five types. We photocopied representative cases of these five types and included them in Chapter Two.

Further theoretical sampling led us into collection and analysis of written materials of test critics, testing professionals, and the public. For each group, we identified unique and contrasting properties of the core and other categories. When these were distinguished, we searched for excerpts that would best illuminate these beliefs. To accomplish this efficiently, we reconstructed portions of written text from newspapers, newsletters, public testimony, and the like into a form that seems to have one voice. The unreconstructed and reconstructed texts were submitted to an outside expert to verify that we had not misconstrued the beliefs as revealed in the original documents.

Other data, such as the final interviews with focal teachers, the group interview over the ITBS test booklet, statements recorded during observations of classrooms and meetings, the content analysis of newspaper articles, newsletters of FAIRTEST, and other documents, were used to generate further properties of the core category and suggest theoretical interpretations, apropos beliefs, about testing and how they are organized. These alternative methods of collecting data permitted us to triangulate assertions and concept definitions, and circumvent errors and distortions that come from results based on single methods.

Analysis of observation data. We followed the guidance of Erickson (1986) to analyze the data from observations and documents. For him, qualitative data analysis consists of generating assertions from data and presenting the evidentiary warrant for those assertions.

Taking the data record as a whole, we proceeded inductively to generate assertions (analytic generalizations or general conclusions) from the data. We read the data record repeatedly, searching for themes or principles that could organize the data and answer the questions of the study. Working hypotheses, observers' interpretive comments, and categories which we had discovered in the analysis of interviews served as templates for this analysis.

The assertions, which are presented in Chapter Four, vary in the degree of inference from the data. For example, consider the assertion that reads as follows: "The role of testing changes over time in relation to the schedule of external tests and the time of year. A natural history of the testing event serves to organize participant actions and meanings with respect to testing. Actions and meanings of teachers and others change through the year in recognizable stages before, during, and after the test and the publication of test results." This assertion requires a low degree of inference, and the reader can readily follow the train of logic.

To test the "natural history" assertion, we constructed a two-dimensional display (Miles & Huberman, 1984) with time (over 15 months) on one dimension and role of testing (e.g., preparing, testing, resting) on the other. We recorded the presence or extent of test-related activities in each cell of the matrix. Finding few disconfirming instances—activities off the main diagonal in the matrix—increased our confidence in the assertion. To illustrate the cyclical nature of the testing event, we searched the data record further for significant excerpts. These we converted into vignettes. The purposes of vignettes in qualitative research reports (Erickson, 1986) are (a) to provide "particular description" that convinces the reader that the researchers were in sufficient physical and psychological contact with the participants so that their meanings and actions could be understood and portrayed; and (b) to illustrate the data that led the researchers to the assertions. These vignettes are incorporated into Chapter Three.

After the initial draft of the "natural history" was written, it was submitted to six participants in the study, including the focal teachers and principals. Three of the six asked that modifications be made in the tone or content of the narratives that they felt reflected adversely on them as professionals or compromised their anonymity. The requested changes made no issue of the adequacy of the "natural history." Six experts in either qualitative research or testing programs read and reacted to the draft, leaving the interpretations intact. Thus, this assertion stands the following tests: rigorous search of the data record for disconfirming instances (Erickson, 1986), internal reliability through verification of multiple observers (LeCompte & Goetz, 1982), verification by member checks (Miles & Huberman, 1984), reader credibility (Eisner, 1981), and verisimilitude (Phillips, 1989). Other assertions are illustrated with vignettes and general description in that chapter. These assertions also met the assay of evidentiary warrant named above.

In the final stage of analysis, we reviewed the professional literature on testing and compared the evidence of the present study to available theories and evidence.

The remainder of this report is organized in the following way. Chapter Two presents the analysis of beliefs about testing. Significant excerpts are included to reveal to the reader the details of the analytic process leading from data to theoretical assertions. Chapter Three, entitled "The Natural History of the Testing Event," presents data and analysis about the organization of activities related to testing at the two schools. Chapter Four covers additional assertions about the role of external testing and the impacts of testing, and presents an overall interpretive framework for the study.

Chapter Two: Beliefs about Testing

To understand external testing in elementary schools, one must study how school people define and interpret its meaning. The same standardized achievement test given in two schools may carry quite different meanings; one teacher may define it as a week's respite from reading lessons, and another define it as a threat to her freedom to teach as she chooses or a threat to teach at all. According to the conceptual framework that guides this research, persons' meanings and definitions of situations are sensitive to social contexts and in turn influence the role of external testing.

One aspect of studying participants' meanings is to describe and analyze their verbal discourse: what they say or write about testing. Beliefs, or verbal statements about what one holds to be true (Price, 1969), are the articulated and conscious part of participant meaning. Beliefs may be held with greater or lesser justification or warrant, but they fall short of ideal knowledge. Knowledge, as distinguished from belief, is public, agreed upon by consensus, and well-substantiated in logic and evidence. In this research, we consider "knowledge" about testing to be evidence in the literature of psychometrics. Among most of the participants in this study, knowledge in this sense is scanty, but beliefs about testing are plentiful. Like most teachers, these generally lack psychometric knowledge. Few, for example, would be able to state in propositional form that "Twenty-four percent of the variance of achievement test scores can be accounted for by family socioeconomic status"[1]. However, teachers have personal knowledge about testing that comes from direct experience with specific pupils, events, and circumstances. An example of personal knowledge might be a teacher's repeated observation in giving the ITBS that some pupils give up on portions of the tests that they consider too difficult, even when the true educational attainment of these children is appropriate for their grade level. This kind of personal knowledge comes from direct experience rather than textbooks and lectures by testing experts.

To study beliefs about testing, we conducted formal interviews, recorded statements teachers and others made about testing during our observations of school life, and collected statements made in public records. In this chapter we present our analysis of the beliefs about testing. In support of this analysis we describe and interpret teachers' beliefs and contrast them with beliefs expressed by school and central district administrators, pupils, the public, testing experts, and critics.

Teachers' Beliefs about Testing

Formal interviews with teachers at Jackson and Hamilton Elementary Schools covered four primary topics: what information teachers glean from test results, how they use test scores, effects of testing they perceive, and test preparation. Inductive, grounded theory analysis of the interview transcripts yielded 63 categories concerning beliefs about testing. As Table 1 illustrates, these categories cover concepts and mental processes such as "Defining a valid test as one constructed by someone familiar with local curriculum and circumstances," "Relating the error of testing to the day-to-day characteristics and idiosyncrasies of pupils," "Using tests to justify special placements," and "Specifying the opportunity costs of tests and test preparation," as well as topics such as "Describing the characteristics of local pupils."

---

[1] Mayeske (1973) cited in Berk (1988).

Using constant comparative analysis helped us define these categories and hypothesize about the relationships among them. The category that occurred most often in the analysis, helped to explain most of the data, and related to the greatest number of other categories was the following: "Defining the discrepancies between the indicator and the trait of achievement." In this category were issues such as (a) how teachers define educational attainment and educational goals, (b) what teachers believe about what the achievement test scores indicate, (c) how the indicator represents only a small part of what teachers define as attainment, (d) how the indicator or score is distorted because of differences between what is measured and what has been taught, (e) how poor pupil ability, motivation, language, social class, and the like confound test performance and increase the discrepancy between the indicator and the true level of educational attainment, and (f) how technical features of the test detract from test scores. Combining these limiting and distorting influences on test scores, one concludes that teachers believe tests to be poor reflections of what pupils accomplish in schools.

We compared each instance of a category with every other one to generate its properties. The aim was to discover all possible properties of a category. For the category, "Defining educational attainment," we coded all instances of teachers' statements relating to this idea, identifying 23 nonredundant properties—qualitatively distinct ways teachers defined educational attainment. Some of these were further sorted into two subcategories, DEFINING ALTERNATIVE, UNTESTED GOALS, and DEFINING EDUCATIONAL ACHIEVEMENT. Properties of the latter fell into a dimension of "Consistent or Inconsistent with achievement testing models." Different teachers named different properties of these categories. Some were named by one teacher and some by several. It is not the goal of grounded theory analysis to provide statements such as, "Forty percent of the teachers gave 'basic skills' as their definition of achievement." Rather, grounded theory analysis aims to explore the possible meanings and qualities of the categories and understand their interrelationships.

Teachers Define Educational Attainment

How do teachers define educational attainment? Teachers in this study seem to define educational attainment not solely as products or bits of knowledge and skill that result from teaching, but also as social and psychological processes. When they reflect on what they consider their teaching responsibilities, teachers are more apt to name processes such as "helping kids develop an understanding of multiplication concepts" than outcomes such as "pupils can attain correct solution of 9 out of 10 two-digit multiplication problems." In this respect, teachers' beliefs are sometimes inconsistent with existing models of achievement testing. According to Resnick and

Table 1

Lists of Categories and Codes from the First Stage of Analysis of Teacher Interviews

Validity Codes

VT1	Defining what it is that achievement tests indicate
VT2	Defining a valid test as one that measures district curriculum
VT3	Defining a valid test as one constructed by someone familiar with local curriculum and circumstances
VT4	Making grade level interpretations of norm-referenced tests
VT5	Defining the discrepancies between the test requirements and characteristics of local pupils
VT6	Defining a valid test as one that surveys knowledge broadly
VT7	Asserting that emotions, intentions, and motivations influence test results and increase the discrepancy between the indicator and the trait
VT8	Exploring the discrepancies between the teacher's judgment of pupil achievement (one indicator) and test scores as indicators of achievement
VT9	Defining the discrepancy between performing a skill and accurately answering test questions that purport to measure a skill
VT10	Comparing the validity of a teacher-made test (a competing indicator) with the external test as an indicator of achievement
VT11	Defining what a test indicates is only what the pupil knows or can do on a particular day
VT12	Defining a valid test as one that measures the trait named in its title
VTCONTENT	Defining validity as the extent to which a test measures what has been taught
VT=TEXT	Defining a valid test as one that measures what the textbook covers
VTPROG	Defining the validity of a test for assessing the effectiveness of a program

VTTEACH	Defining the validity of a test for assessing the effectiveness of a teacher
VTPLACE	Exploring the validity of tests that are used for special placements
VT=TRUTH	Associating validity with truth or indicators with "true" scores
VT=FAIR	Associating validity with fairness
VT=RT	Associating validity with accuracy or reliability or the absence of chance in obtaining the score
VTTS=OS	Defining the discrepancies between the indicator and the trait of achievement

Reliability Codes

ITEM	Arguing the merits of particular test items as accurate, fair, and true components of indicators
RTLUCK	Describing testing events in terms of guessing and other elements of chance
RTTECH	Distrusting the technical aspects of test administration, scoring, norming, and reporting
RTTEMP	Relating the error of testing to the day-to-day characteristics and idiosyncrasies of pupils
RTSAMPLE	Relating the error of testing to inadequate sampling of content from the content universe
RT1	Relating the error of testing to inadequate tapping of performance (single vs. multiple assessments)
FORMAT	Relating the error of testing to tricky or unfamiliar formats employed by the tests

Utility Codes

UT1	Defining the utility of tests in relation to teacher assessments
UT2	Defining the utility of tests in relation to external audiences (accountability)
UT3	Defining the utility of tests for teachers given the scope and timing of the testing event
UT4	Defining the utility of tests in relation to the total testing program (redundancy)

UT5	Using tests to justify special placements
UT6	Defining the utility of tests as a tool for managing instruction
UT7	Defining the utility of tests to train pupils in test-taking skills
RITUAL	Ritualizing testing of low utility, low validity tests
Effects codes	
EFFS1	Defining the effects of tests on pupils (duration, frequency, difficulty)
EFFS2	Asserting the relationship between anxiety and other emotions and the effects of tests on pupils
EFFS3	Arguing that effects of testing depends on the type of pupil
EFFS5	Arguing that effects of testing depend on teachers' attitudes about tests
EFFS6	Arguing that tests detract from classroom community and cooperative learning
EFFT1	Defining the effects of tests and test scores on teachers and school principals
EFFCURR	Defining the effects of tests on curriculum and teaching methods
Preparation Codes	
NATHISTAE	Describing the natural history of the testing event from early in the school year (A) to the week of the test itself (E) and afterward (F)
PREPCOST	Specifying the opportunity costs of tests and test preparation
PREPFORM	Describing the process of preparing pupils for the formats of external tests
PREP1-5 PREPCUE PREPBST PREPSST PREPITBS	Specifying the preparation for various tests
Other codes	
ALTER	Specifying the conditions under which standardized test administration procedures should be altered

AVERAGE	Expressing concern for the class or school average of test scores
03	Defining the discrepancy between the school's program and district curriculum and the meaning of "external tests"
AUTONOMY	Defining issues of teacher autonomy and authority in relation to tests
CHARPUPIL	Describing the characteristics of local pupils
CARING	Expressing emotions and describing actions expressive of caring for pupils in nonacademic ways
ACHVMNT	Defining the characteristics of achievement in relation to testing of achievement
REACTION	Expressing emotional and cognitive reactions to existing test scores
ALT-GOAL	Asserting alternative and untested goals for schools
PRESSPASS	Describing the pressures put on teachers and others to raise test scores
ALIGN	Describing the alignment of instruction to match test content, format, or sequence
PARENT	Describing the communication about and interactions with parents about test scores
COMPARE	Defining the processes and effects of comparing test scores across pupils, grade levels, schools, districts
LEVELS	Describing the processes and results by which pupils are tracked or placed
IMAGE	Defining the image of pupil, teaching process, learning
KNOW	Knowing (tacit) and being told (propositional knowledge from tests)

Resnick (1989)[2], models of achievement testing assume that important educational attainment can be adequately represented as concrete products of children's behavior—skills and bits of knowledge that result from teaching.

Instead, teachers often define attainment as processes such as "developing productive relationships among kids," or instilling "a sense of community in the classroom," "improving the dialogue between teacher and pupil"; and promoting "the growth and change of the whole person," the "ability to make choices" and "assume responsibility," "understand the structure of and have a feel for literature," and "become committed to a task." They name as important goals for themselves the instilling in pupils of "productive strategies of reading and learning," "loving books," and "electing to read," "expressing themselves," committing to "the quality of the things they do," and "experiencing the spirit and excitement of learning."

Defining achievement itself as a subset of their goals for education, the teachers name such properties as the "ability to do abstract kinds of problems," skills that the pupils "can do at the end of the year that they couldn't do at the beginning", "thinking skills," "logical, creative, reflective thinking style," "understanding how to attack story problems, not just the ability to work the math problems," "the ability to find spelling and grammatical errors in their own writing," and "Something more than what is on the test!" Other properties of the teachers' definition of achievement are "all of what pupils should know to be well-rounded"; ability to "write out whole sentences rather than just filling in the blank"; and "putting out and trying hard on a daily basis." Although they refer to the products of teaching, the latter properties of teachers' definitions of achievement fall outside the achievement testing domain.

Some of the teachers define achievement in terms more consistent with achievement testing models; for example, as that which pupils "know on a particular day"; or as "basic skills—you keep reviewing it, and it becomes part of their learning...what they've retained"; "growth from year to year on what the tests measure"; "perfect mastery" of what has been taught; "meeting an objective"; "long-term retention" and memory of what has been taught; or "scoring high on tests most of the time."

Hamilton teachers tended more often than Jackson's to hold views of attainment and achievement more consistent with achievement testing models. However, the correlation, is imperfect. This is one of the few categories in which beliefs are organized by school. In other cases, beliefs about testing are shared by the faculties.

---

[2] Resnick and Resnick (1989) argued that the model of learning embedded in achievement tests has the following characteristics: assuming that knowledge and skill can be "decomposed" into independent, additive components, the sum of which indicates the knowledge and skill as a whole; decontextual performance, that is, assuming that "each component of a complex skill is fixed, and that it will take the same form no matter where it is used" (p. 11), or on which task it was originally based; learning is that which results in correct responses defined as such by someone else and in advance; assessment of learning is governed by technical considerations such as reliability, at low cost per unit of information; and judgment of responses can be made by a disinterested third party.

Teachers Define What They Believe Achievement Tests Show

Compared with their definitions of the real underlying trait of achievement and attainment, consider what the teachers believe achievement tests show. Some say that the ITBS "shows growth from year to year," although they believe that what is "growing" is as much a matter of growing test-taking abilities, effort, attitude, and perseverance as it is achievement. On whatever it measures, some teachers believe that tests like the ITBS provide a basis for comparing their pupils to pupils across the nation. Some teachers believe that tests such as ITBS and CUES help them "identify those pupils who have not mastered a concept." Achievement test scores "reflect how pupils do on worksheets," "tell what a child knows on that one day the test is taken," sometimes "tell who the smart pupils or good students are," or "sometimes confirm what is already known about their achievement." Some teachers say the scores help them "identify problems or abilities of pupils with extremely high or low scores" or provide a "guideline, a place to start planning instruction" based on information about "gross" ranges of achievement. Some teachers believe that test scores tell only how well teachers covered material on the test or faithfully followed the program on which test is based. Most teachers are likely to believe that achievement tests in math are consistent with their definitions of math achievement, or at least the part of math achievement that covers "math facts."

These properties of the category, "Defining what achievement tests indicate," show that teachers in this study are far from being committed antagonists of achievement testing. Most believe that tests such as ITBS, BST, and CUES measure educational attainment, for some pupils and certain circumstances and for particular, but limited, meanings of achievement. As one teacher said about the ITBS, "Yes, it does show something about achievement. It has to."

Neither are the teachers confirmed apostles. Each acknowledges some degree of discrepancy between educational attainment in the ideal sense and the achievement test score. Quoting one primary grade teacher about the ITBS, "At least not at this level does it tell me anything. I don't believe I get any information from it that I don't know already. I guess it can sometimes confirm some things I know about children."

Although some of teachers define educational achievement in terms consistent with educational testing models (i.e., as outputs of learning), all the teachers in this study profess the belief that achievement tests reflect only a diminished and perhaps skewed portion of the goals for which schools strive.

Teachers Define the Discrepancy Between Test Scores and Educational Attainment

Studying interviews and statements recorded in the everyday activities of teachers, we delved further into their beliefs about the discrepancy between the standardized indicators of achievement and their own definitions of achievement and attainment. Is the discrepancy one of kind or degree? Is it greater under certain circumstances? What are teachers' beliefs about its causes and effects? Properties of the category DEFINING THE DISCREPANCY BETWEEN THE INDICATOR AND TRAIT OF ACHIEVEMENT include characteristics of pupils, curriculum, the tests themselves, and the social and educational contexts in which they are given.

Features of pupils. Teachers believe that the discrepancy between the indicator and the trait of achievement is more pronounced for certain types of pupil. The discrepancy is less for pupils of above average intelligence. For this group, other things being equal, the score on the achievement test may approximate a pupil's true achievement. But tests may bore very bright, creative or divergent thinkers or such pupils that "read too much into test items" and choose the wrong answer. Many teachers believe that tests such as the ITBS measure intellectual ability rather than achievement, and those pupils with below-average intellectual ability may score lower than their true achievement. Pupils' ages also figures into teachers' estimates of the discrepancy between the indicator and trait of achievement. By sixth grade, pupils can read well enough and are accustomed to achievement testing, so their scores are more apt to approximate true achievement than are those of primary grade pupils.

Children also differ in their emotional stability, self-confidence, and motivation, and teachers believe that achievement test scores reflect these traits as much as they reveal true achievement. Therefore, we conclude that teachers believe that the discrepancy between the indicator and the trait of achievement is less for children who are emotionally stable, confident, and motivated. For other children who are having trouble at home or with friends or whose parents have neglected to instill in them the habit of perseverance, the test score fails to reflect their real attainment. The following are characteristic instances of these properties.

> Sometimes [the scores are] right on, sometimes they're way above what they should be, and sometimes they're much lower than what they should be. You know there's something that depends on the day.
>
> I know the smart kids. You know them. And they're the ones who are scoring high on the tests. So there is something about that it's telling us.
>
> If I were to read a question over again and think about that child and what he knows, his knowledge, where he's at experientially, then I could better understand why he missed a question. So I need to know how to interpret what he missed.
>
> I started out by saying how important attitude was. So if a student comes to this testing thinking this looks easy, it more than likely will be easy and the student will do well. If even as the papers are passed out, the student has just bombed out or been chewed out, he may be frustrated even before he opens the cover, and it will show it differently.
>
> I think the test is very much geared to a person's ability to read. Because the sections that have to do with maps, or some of the sections that are on the test in math, you have to be able to read the words properly in order to answer the questions correctly. And if you can't read it right, I think it would be very, very difficult to answer a question where you really could do the computation if the words had been a little easier to read.
>
> Mostly I look at what the child does in the classroom. And many times I just kind of chalk it [low test performance] up to being emotional, and I just say,

"Well, you know, this child has an emotional problem but he can still learn and he's going to."

And so the scores may not indicate anything other than, "I just don't care to make that effort today." Or there was a problem at home. And there are so many problems with so many here. I have one here who sits right there, nice looking boy and at the beginning of the year he was so up and positive and sure of himself and doing well...And now there's a breakup in the home. And he's absolutely, you know, demoralized....So those tests, you know, everything has to be perfect and ideal. I mean everything has to be right. They have to be willing to overcome all obstacles that may be bothering them emotionally and they have to push the button and say, "I'm going to do my very best and do it." Then maybe the test will show something.

Everybody is lumped together, and a child that has an IQ of 75, that we can't get tested [for special education] or a kid that's holding on by his fingernails to even get to school and make it through the day, is taking the same test alongside a kid with a 140 IQ, and the teacher is responsible for both. Well, if the 140 IQ is spacing off, because he's bored, or thinks it's stupid to drop dots, and the one with the 90 IQ is freaking out because he can't read the test, and one or the other or both of them scores poorly, it's the teacher who's [held] responsible. And I don't see the correlation.

[What makes the kids not care is] their background. The home. They don't see any reason for it. They don't really have any support or that kind of thing, or they have too much, and if they have a bad grade they are punished for it. That sort of thing, instead of supported or helped. I think it's how parents react to the scores.

You would have to take that test with about a fourth grade mentality....You cannot deviate from what some teacher would have pounded into your head as rote learning.

Well, I think some of the kids are more attuned to tests than other kids. I think we have to learn to take a test. I think test-taking is a skill and I think there are students who are tuned into how to take a test. They know how to make a good guess. They know how to choose. They know how to knock out information that is not relevant and they know how to then make a choice, whether that's a study skill or whether they've been taught to do that.

You have two extremes. You have the kid that thinks it's a big deal and kind of wants to do well and gets tied up, they don't perform worth a toot on the test. And the next day they sit there and tell you reams of information that couldn't pull out to save their souls on the test. Then you get the other kind, that just sit there and make little dots and they don't care. They don't even read the questions. You know. They're through with a section in five or ten minutes. It takes any thinking person twenty five to thirty minutes to do. They're through with it. They don't care. It means nothing. It doesn't mean anything at home, so it doesn't mean anything to them.

> There are definitely some children who as soon as you say test, they freeze. And they are not good test-takers and they score low. And then, especially, if you are giving some 11 tests in the spring of the year, I think they get to the point where they're almost paranoid about having to take a test again. And I think we're putting a lot of stress on the children because of that.

Features of curriculum and instruction. By this analysis, achievement tests distort educational attainment because their content and format rarely reflect what has been taught in the classroom, particularly if the educational processes and goals diverge from the educational testing models. For example, many teachers aim for "authentic literacy" in their pupils, and attempt to achieve it by organizing opportunities for the pupils to read works of literature and engage in authentic opportunities to communicate orally and in writing. Teachers believe these activities and goals are out of step with the kinds of skill that achievement tests cover. Similarly, some primary grade teachers aim for conceptual understanding in their pupils by having them manipulate concrete materials and count and arrange objects, an endeavor quite different from "the skills approach" to arithmetic that the achievement tests cover. Teachers who emphasize cognitive or higher order objectives express frustration with the "rote-memory" or other "low-level objectives" that tests emphasize. These teachers believe that the scores obtained underestimate their pupils' true level of attainment and the teachers' true accomplishments.

> What we really care about here is how kids are reading real books and writing and understanding. But the Iowa tests component skills that don't necessarily add up to anything.
>
> It [the achievement test score] tells you exactly what it is supposed to tell you, that that child on that particular day can pass that skill. It does not mean that he will know it forever. It doesn't mean that he'll know it the next day. They were designed to tell you that at the time they took it they knew it. And that's all it does.
>
> It's not a true evaluation of what they do and what they know. Some, I would say, yes, maybe so, somewhat. But it isn't actually testing on what they're learning and how far they've come, I don't think, except for how they do on worksheets. So for some it might be a valid assessment of what they know....or how well they can do on paper. And I'm not talking about writing or applying it to life, I'm talking about how well they do a worksheet.
>
> I know what I've taught. I know how I've taught it. And I know that if there's any form[at] involved I've given it to them. And I feel like then I can create a test that would be honest. The kids would not be surprised by the form[at]. They'd not be surprised by the words—I mean the style. They'd not be surprised by the content....(ITBS), I suppose that it does tell something about children's achievement. But it doesn't test necessarily what's taught. I think that's the thing that would bother me more than anything else...Someone has set up the test who has nothing to do with our district, and so they may be testing things that we haven't taught.
>
> When you teach children in this way you teach them that there is more than one answer to problems. So they out figure this test horrendously. And

> their reasons are very good if you could listen to their reasons. They're right in my opinion, even though their answer is wrong....Because we've taught them that it's all right [to question the questions], and you have the power to say something that I may not agree with, but it's your right to think it. [But I don't say that in preparation for the test], I'd mess them up too much.

Features of tests. Features of the tests themselves also contribute to the discrepancy between the indicator and the trait of achievement. Teachers know that to avoid a ceiling effect at any one grade level, items on the ITBS must span a range of difficulty, with some items being absurdly difficult (e.g., fractions on a test made for second graders). Necessary though the difficulty may be, it causes some pupils to give up before they can demonstrate their actual achievement. Teachers understand that standardized achievement tests must be long enough to be reliable and must sample from a universe of content, yet they believe that the length of the test detracts from pupils' performance and increases the discrepancy between the indicator and the trait of achievement. Other things being equal, standardized test items show better discrimination the closer their difficulty approaches 50 percent. In mastery programs, however, pupils commonly encounter items that can be answered correctly at rates of 100 percent. Teachers believe that children become frustrated and confused with the difference and perform at less than their best.

In addition, teachers believe that the multiple choice format limits the range of possible educational goals to those that can be easily tested, a problem that characterizes both norm-referenced and criterion-referenced tests. Teachers cite many instances when pupils guessed and obtained good scores, even when they had learned little.

For pupils accustomed to working in groups or getting help from the teacher, the test presents a foreign and restrictive environment, contributing to scores that are lower than they should be. More than anything, teachers resent the ambiguous and foreign wording that standardized tests of achievement use.

Beyond the characteristics of the separate tests, teachers believe that the many tests in the testing program exhaust the pupils and eventually lower their performance on tests taken later on. Here are some characteristic excerpts of these and other relevant properties of this category.

> It's limited because it only tests certain things. It's limited strictly to skills.
>
> Group questions, would be, you know, lower level questions.
>
> The third day, you could tell they were tired. And by Thursday, I was looking at the class, where kids had peanut butter for brains. They were absolutely whipped. And the information that I got off of them, as far as a test score, was just a number. As far as I was concerned, it meant an endurance score. It didn't mean anything.
>
> I don't think you should start with one of the harder things in the [Iowa test] book. Kind of blows them away. And we've done that a couple of years. We've asked them to do what I feel are the two hardest subtests in the tests first and then, I guess in some ways you could tell them, "Okay, the rest of it's much easier than this. So get geared up and try again." But for some others,

they're gone. That's it. You've blown them away on the first two pages, they don't know what you're talking about, so forget the rest. Some of them are hard to change their attitudes, too. Once they settle into, "I can't do this," and do any old thing. And others of them, you see, especially on math, because it's so scattered, they'll do an easy one, then a harder one, then later on there will be a real easy one. So if you can get them to really look at it and look at each one differently, and not just give up on the whole thing.

By the last test (BST-Social Studies) the kids just didn't care. I could see it. And I know and feel certain that the reflections didn't come from me because that's my baby and I was so excited about it. I could see it, and they did poorly, very poorly.

With the Iowa, the child has no choice about the task. They have no choice about how long they sit there. And there is always the opportunity for them to just mark just to get it done and be finished with it. And that mark, whether it was put down for no purpose whatsoever, or whether it was their best attempt counts just the same.

I've had kids I know were good students, who should have done real well, who just couldn't do it. It was just psychologically so overwhelming for them that they couldn't apply themselves and do the best job that they could. And those kids just kind of, you watch them crumble. I think the tests are, many parts of it, too long. They ask for much more attention, particularly the reading. It's just a monstrous amount of pages turning. They look and see if they have to turn the page again. It's "My God, we just did two whole pages and all this reading and all these questions." And that's not what their daily work is like. So it's such foreign activity to them and they're just not at this age geared to doing that.

I had some of my top students go back two years [attained scores two years below grade level]. I mean lost. They unlearned two years. My very best students. Then I had some who increased by two years. I mean they just hit the casino at the right time, pulled the lever, and they got it. I mean they just scratched out those right answers. Well, it has to be something like that. You come to the end of the test and there's ten minutes to go and so you just blank in those little dots. And that's not testing achievement.

But as far as them being not valid, when a child sits there and guesses at all of the answers, obviously it's not going to be valid whether he scores high or low.

When I've seen kids that I have eyeballed a subtest, and know for guaranteed fact that they had a perfect score, and see that come back with errors on it. And when I get others that I know that there were a bunch that were wrong, and you get a subtest for a whole section is wrong, and you get a printout that says they're all right, it really makes it suspect of the validity of any of the scores.

I think children are naturally geared for that. They're geared for "100 percent is best." And if you don't get 100 percent that's not the best. I've seen kids throw papers away because they were 96 percent and ashamed to take them home. They're so used to getting 100 percent on something like that. So you give them a test and then naturally they think they have to answer all the questions. If they don't answer all the question, they've done poorly. You know. and that's the way it's been geared all along to suddenly give them a test and say, "Well, it's all right if you don't answer that." I don't think they can comprehend that. That they can still do well on a test without answering all the questions.

They word questions to deliberately frustrate them. I saw an example the other day. They are to find the words with the long "a" sound. In all of the examples, all of the choices are irregulars and there's not an "a" in any of the words.

There was a picture and there were three things that belonged on a farm, and one was...there was something else that didn't belong on a farm. I can't remember what it was, but this child picked the two things that belong on a farm and something else. And I said, "Why did you pick that?" And he said, "Well, because all three have wheels." So he was operating from a different concept altogether than what the test maker had in mind.

Okay, you have to understand how first graders decode words. First of all they look at the picture. Then they say what they think it is. Then they look for the word that looks closest to that word. If they look at this picture and say it's a light, which most of them would, and then they look over here [to the options] and they don't see anything that begins with "l." They sit back and look dumbfounded, not being able to read all these in isolation, they don't have any strategies to go back.

When we get ready for the Iowa, the thing that is so fascinating is that your best thinkers, your very best students, your most creative and original children look at those sample tests that you can talk about, and they'll say, "But, three answers are right." You know, that kind of thing. And give you wonderful reasons why they're all right. And you're left to say, "Yeah, good thinking. Wonderful thinking. It will get you the wrong answer on the test. Quit thinking, It will not help you on the test."

I see it over and over. I see the bright kids having trouble with those tests because they haven't learned just to put on the blinders and just go right down the road and don't think. Just do it, just do it! Don't try and do your usual creative things. Don't do it! That's not what they're after. So you've got to put the blinders on, you can't think, "Well, what if?" If you think what if, you're shot down. Because the test makers aren't thinking, what if. They're just plodding along in the same little rut, and you have to do the same thing, to take the test.

The CUES are just so ridiculously easy, they have to be easy so the kids can make 90 percent the passing criterion. Then what happens is that they pass the CUES but not the Basic Skills, which is supposed to also be based on the Scope and Sequence. But there's no correlation.

Table 2

Defining the Discrepancy Between the Indicator and the Trait of Achievement: Summary of Properties

"Indicator is Discrepant from Trait of Achievement"

When teacher disconfirms indicator with different indicators or teacher judgment.

When what is tested fails to cover what is taught in class.

When the test is too difficult or long.

When pupils guess in multiple-choice format.

Because of technical features of tests (norming, ceiling effects, etc.).

Because test is a single, simple indicator but achievement is a complex trait requiring multiple indicators.

Because a test score fails to measure long-term retention.

Because of the large number of tests in testing programs.

Because items are confusing and ambiguous and require that pupils adopt a particular frame of reference.

When programs teach that there is rarely one correct answer to complex problems, yet tests require one.

When knowledge in minds cannot be translated to test format.

Because achievement tests really measure endurance, diligence, persistence and attitude, IQ, intention.

Because tests of subjects other than reading require too much reading ability.

For curricula other than worksheet, paper-and-pencil.

For definitions of achievement other than short-term low cognitive level.

For the youngest pupils.

For pupils who are divergent thinkers, who read too much into the items and answer incorrectly.

For pupils who do not know the importance of testing.

For pupils who are frightened of tests and freeze up.

For pupils with low self-confidence, who have failed before and expect to fail on the test, who fear taking risks.

For pupils with emotional or family disturbance.

For pupils with short attention spans, learning disabilities, English language deficits, low SES, high transience rates.

For pupils who are not truly involved with taking the test.

For pupils who do not put in appropriate effort during testing.

For pupils who are not good rote memorizers.

For pupils who are not sophisticated in test-taking skills.

For pupils who lack good motor skills (who cannot work fast or transfer answers to separate answer sheets).

Characteristics of the tests themselves, the testing program as a whole, the instructional program, and the pupils all militate against the inference that a score on an achievement test is equivalent to a true level of achievement. For the teachers in this study, the achievement test score has been substantially, but not completely, drained of their meanings of attainment. According to the definitions of the teachers, a low score attained by a particular pupil or an average of scores attained by the class on any of the external tests does not mean the pupil or the group failed to learn. Some admit that they would feel some satisfaction and relief if their classes attained high scores, yet this feeling would be qualified by the realization that the kind of learning that the tests measure is not long retained or often applied. Some teachers attribute high scores to luck, pupil intellectual ability, pupils' having learned test-taking skills and the like. To teachers, achievement test scores, whether norm- or criterion-referenced, can only be meaningful if they can interpret them in light of other indicators, their personal knowledge of how hard the pupils worked on the test and what other things were going on in pupils' lives, the characteristics pupils bring to the test, and the match of local curriculum and contents of the tests.

Teachers Define the Utility of Test Results

Aware of the distinction between the indicator and the trait of achievement, one would logically expect teachers to moderate the place of external tests and test scores in school life. Consistent with their definitions of the discrepancy between indicator and construct, the teachers claim to make almost no use at all of the results of external tests. They rarely use ITBS, BST, or CUES in planning instruction or grading classroom performance even though the latter two purport to follow district curricula. They rarely examine ITBS results except in those instances where children's classroom performance is extreme or anomalous. Teachers sometimes bring ITBS and CUES results into discussions of special placements, such as changing pupils' reading group of, retaining them an extra year in a grade, or referring them for special education. However, they use the scores to support other indicators of classroom performance and ability that are more meaningful to them. Teachers also use test scores to defend their professional judgments, when those judgments by themselves are not credible. The ITBS has particularly low utility for teachers because it is given in the spring of the year, and its results are not available until the last week of school. The pupil's teacher in the following year has no firsthand knowledge of the effort the pupil expended on the test itself and so accords it little importance.

The teachers in this study hold opinions about the utility of tests that are similar to teachers surveyed in other research (Dorr-Bremme & Herman, 1986). Although they do not use the scores themselves, they believe that someone else does. They believe that administrators, parents, board members, state officials, and critics of schools use the scores against them.

Teachers Define the Effects of Testing

For pupils, particularly younger ones, most teachers believe that standardized testing is "cruel and unusual punishment." In this section we summarize and illustrate the properties of the category, DEFINING EFFECTS OF TESTING ON PUPILS.

The "length of tests," their degree of "difficulty," the "number of tests" to be taken, the "time limits" involved, the "lack of choice" in the administration of tests, the "individualistic" nature of test-taking, the "fine print in tests and answer sheets," and "the difficulty in transferring answers to answer sheets" are believed to produce "stress," "frustration," "burnout," "fatigue," "physical illness," "misbehavior and fighting," and to be "psychologically overwhelming." Some teachers believe that the "tests cause test anxiety" that would show up on later tests, and "set up a failure mentality." "To get the pupils to maintain effort on the tests," teachers believe that they must "promise treats, rewards, and breaks from work." Or if the "fear and frustration become too extreme, pupils must be told just to do their best and not worry about performing." In extreme cases, some teachers feel that it is in the best interests of learning disabled and emotionally disturbed children to stay home during the week of testing. Teachers believe many pupils simply "give up trying to perform when they encounter items that are too difficult for them." Pupils "worry that test scores determine their course grades or promotion."

> We know that our class has to score. We've got to show growth in the scores they have. And how do we convey it to the children that there is pressure? I think just through our attitude. Children can sense it. They know when there is pressure being put on them when teachers are under stress.
>
> They're wiped out. They cry. They're distraught. They can't find their place. We go too fast. They're too tired. They're so frustrated. They break their pencil in half from just the tension. Some of them hold their pencil so tight they literally break it in half. It's just a horrible week. Sometimes it takes me two weeks afterwards to get those kids' confidence back again. They are just devastated. It's so long. It goes on and on. It's like, "All these things that I don't know and all these things I can't do." And they're just devastated.
>
> If a first grader, or any child, comes to a math problem that's beyond them, and that's what is on these tests, you can see from the first mistake they make that they don't understand it. You can see their expressions or the tenseness if you're walking around watching. You can tell if they've hit a place in that test that bothered them. Even if the next question is something that they know, they're going to tense up again and be wary. And first graders who don't really have a lot of experiences meeting some of these obstacles in life....And these children, they just exist. They don't really know how to cope with a setback like that.
>
> I've seen children unable to sound out words [on the test] that they've sounded out for me in reading groups every day. And if they really have a block, I'll put my finger there and say, "Sound it out" They'll say the word, and I'll say, "Okay, go on." If they don't, basically you tell them to guess or you tell them to skip it. That's all I really do. I don't help them with the test. You know you also bribe them with treats and fun things if they've done a good job, if they've worked all the way through it.
>
> If the fear reaches the take-over point, then you should say, "It doesn't matter," or tell yourself, "Who cares."

A few teachers believe that tests have "no effects on pupils because pupils do not care how well they perform," or that "by fifth and sixth grade, pupils have learned how to react to tests so that effects are less." A few believe that "effects on pupils depend on how the teachers handle the testing situation."

> I wish they [sixth graders] would get a little bit more uptight. No, it's really something that's of no value or much importance at all. Zilch. There is no impact whatsoever.
>
> I try to convey the attitude to the class that says, "Let's show people how smart we are." I feel that my attitude about the test carries through the whole class from day one. And I have heard teachers talk in the lounge about how so and so fell apart. "I hate this test." And I feel that because of the atmosphere of the class, the children are uptight about it.

Tests also have effects on other teachers. Although "teachers' jobs do not depend on attaining high achievement test scores," "test scores affect some principals' job tenure, evaluations, or promotions, and they translate that pressure onto teachers." Teachers experience "shame," "embarrassment," "decreased self-esteem," and "pressure" when scores are low. "Low scores result in flack from parents." "Experienced teachers have learned to rationalize low scores" in terms of pupil ability and discrepancies between the test and the curriculum. "Although principals interpret test scores in light of the ability and backgrounds of the pupils, district administrators and the public do not." "Teachers feel that outsiders associate and identify the teachers' efforts with the scores their pupils attain," which results in "finger-pointing," "blame," and the conclusion by these audiences that the "teacher with low scores has not worked hard." Teachers are aware of the "contest or competition among schools to attain the highest scores" or to avoid the lowest scores. "Low test scores are used against teachers to make them teach in ways they do not choose." For teachers in low-scoring schools, "gain scores are easier to accept than absolute performance reflected in percentiles or grade equivalent scores." A few teachers perceive that, in their colleagues, test scores begin to replace professional judgment in the selection, emphasis, and sequence of content.

> The first year I used Math Their Way I was teaching a second grade class and they scored at grade level. But the other second grades in the school scored higher than grade level, and I had to do an awful lot of talking before they allowed me to use that program again. It's very hard to start a new program knowing that the Iowa may be used against you.
>
> At the district level I think they look at grade level results. Then some year the teacher has been identified, which really is stressful. So when people look at those scores, they are my scores. And that put a lot of stress on me because even though I was here [in a school with many disadvantaged pupils], I feel like there should have been growth, there should have been this and that on that test. Was I going to be compared to the next teacher? Were people going to think, "She probably didn't do very much in math because look at those math scores. They really should have been better than that." Because it [the score] is a number, and you can do with that number whatever you want to do. And nobody would have come to ask me if, you know, if I had children who were in and out of school four times that year. Or whether I had seven of them who came into class in March. Or whether,

you know, half of my class were nonreaders at the beginning of the year. All it says is my name and the score my kids got. You always have an identity with your school, and people look at the schools' scores and say, "Your school is way down the list. Why are those other schools higher than you? You must really have some rotten teachers there." You know, you get caught up in that, even though you have this feeling that you should be strong, that the test doesn't tell you anything. But I can't help feeling I did something wrong.

It really makes me feel frustrated because I think, "What have I done all year?" I feel inadequate. I feel incompetent. And there's also pressure. There's pressure, pressure—that you're supposed to have these scores. And when you look at these scores, you think, "They're going to think I didn't do anything this year"—administrators, the principal. I think the principals are pushed by the district office, you know. It's all a big contest. Who can do the greatest, who has the smartest kids, who is keeping up with the national average. Because the public, the public is not supportive of the public schools.

I'm a competitor, I want to be up there on top, and it's a struggle. I know my expectations should be high, and I should always be optimistic. But I have to face reality, too. I have some difficult problems and difficult children.

I know of one principal whose job was tied into his test scores and was given a demotion because it was tied into his lower test scores. So I know that these things are out there, and I know that even we teachers put the pressure on each other.

There is probably going to be pressure from the central office. The principal has cashed in a lot of chips to get the school the way it is now, to get the low class sizes and to get to be able to try the program we have. I think they're going to want to see that the children are going to perform.

The only thing I can go on is my own experience from last year when the test results were brought in. I was crushed. I was devastated. And then the principal said, "Now look. I want you to come up with the gains and write them down." So that's what we did. Well, then my attitude changed somewhat because I saw gains.

My contract doesn't state that I am hired to increase ITBS scores by nine months. And I know schools in this district where teachers are being harassed because their kids are only scoring two years above grade level instead of four years above grade level. And when you have a first grader that scores at a third grade level on a test and so the second grade teacher is expected to meet the number of months growth or exceed it, you're putting an awful lot of incredible pressure on a small child and a teacher who doesn't have a prayer of doing that. And here we could not deal with that kind of pressure. Because we would be setting ourselves up for 100 percent total failure, because it's out of our control.

> I think society expects us to have stress put on us to make our kids achieve on tests. For what? For a bunch of numbers that may or may not mean anything.
>
> Teachers come in for a lot of flack. "What have you been teaching all year? It's your fault." Because when the children come from first grade to second grade, the parents at the first conference lots of times will say, "Well, I don't know what he learned last year. He's only ranked on the first grade level and he was in there the whole year." They are implying that the teacher didn't do her job.
>
> To say that the effectiveness of a teacher is based on the outcome of a test is not really fair, because you cannot teach a brick wall. You have to have a willing recipient who will then be willing at the time of the test, and not only be willing, but be able to demonstrate what was taught, and that's not always possible.
>
> I felt confident that they had learned, but in the testing they didn't score as high as I thought they should. I felt that I hadn't presented it in the right way, because I had covered the material, but for some reason it just didn't sink in as far as it should have. And so the next year I probably dwelt more on that area in my teaching.
>
> Last year the third grades did very low in language. And so I got together over the summer with one of the other teachers and we put together a notebook of things that we think should be covered for this year having to do with the ITBS tests. We did try to zero in on language that we thought was definitely in need of improvement and instruction.
>
> When you [change the curriculum to conform to the test], it really stops you from going as far as you can go. Because you say, "Heck, I only need this amount of information for them to pass the test." Whereas if I keep going and going and going, there isn't any limit.

Teachers Describe Preparation for Testing

We questioned teachers about what, if anything they do to prepare their pupils for testing. Most of their responses concerned the ITBS. The activities they claim to engage in are (a) reviewing the content of ordinary instruction, (b) boosting self-confidence, (c) teaching or explaining new material that will appear on the test, and (d) coaching in test formats. Although we were able to classify activities by their intended effects, the teachers themselves could not be so neatly categorized. In this section, we present quotations and interpretations that represent these four classifications.

Reviewing content of ordinary instruction. As one teacher states, "To prepare for the test, I teach what I need to teach." Another professes to reviewing the content of the curriculum that the school prescribes. He claims not to know ahead of time what substance the ITBS includes, and says, "I'm not sure that it will do any good [on the ITBS], but we review and review what the textbooks and the reading programs cover." For him, the texts and the Scope and Sequence are the "higher calling"; his goal is not to produce high scores on the ITBS. Another teacher says her review is not geared for the ITBS but simply to "refresh their memories." She says she tells her pupils, "You've sucked it in. Now it's up to you to remember it." Yet another teacher says she uses the *Math Objectives Review* (MOR) materials, distributed by the district, to review ordinary curriculum, although other teachers use the same materials primarily as a way of coaching test format. Some teachers review most assiduously those parts of their ordinary curriculum that they know the ITBS covers. As one says, "I review every concept that's on there. And if there's something we're having a problem with, well I get as many questions of that type as I can work in. We do worksheets together in class and it gives them some idea of what they can do and what they can't do."

This type of test preparation also includes repetitious practice of skills such as solving math problems orally presented, reading a part of a story and answering questions about it, following instructions with multiple parts, looking at pictures and answering questions about them. Such skills are part of the ordinary curriculum and part of the standardized achievement tests as well.

Boosting confidence. Teachers suggest that they use the materials the district provides (such as MOR) or published materials such as *Scoring High on the ITBS* to fortify the morale of their pupils, get them to believe they can do well, and decrease their anxiety about taking the test. Other means toward these ends include (a) enlisting parents to make sure their children are rested and nourished prior to testing, (b) reminding pupils that their grades and promotion are not tied to their performance on the ITBS, (c) "teaching them the atmosphere" in which the test will be taken, including working alone rather than in groups and not bothering others, (d) "not making a big deal out of testing so it will be less stressful," and (e) indoctrinating them to believe that the test is just an opportunity to show how smart they are or that testing is an unavoidable fact of life. Some teachers intersperse test-preparation activities with "relaxing activities like drawing pictures or making designs," or promise treats and rewards for good effort. Underlying these beliefs is the assumption that teachers need to inoculate their students against the anxieties and frustration of testing.

Teaching/explaining content. Another type of test preparation involves teaching new material that is known to be on the ITBS. Teachers believe that this is necessary when some parts of the ordinary curriculum do not match what is on the ITBS either in substance or in sequence. For example, the Reading Mastery program used at Hamilton uses a special system of print sizes and types and phonetic markings to teach reading skills in early grades. Because the ITBS does not use such conventions, teachers say they find ways of introducing ordinary print and words without such markings. This requires teaching outside the ordinary curriculum. At Hamilton, the language program does not introduce contractions and possessives until about third grade. Because these are bits of knowledge and skills required by the second grade ITBS, teachers claim they invent methods for teaching them outside the ordinary curriculum. One teacher says of the third grade math test, which includes some fractions, that her test preparation includes "giving them exposure to stuff they are really too young for," that is, that are out of the sequence of the ordinary curriculum at that school. Teachers using Math Their Way, a program intended to promote conceptual understanding through manipulation of concrete, manipulable objects, must teach their pupils to translate from the concrete form to the symbolic, paper-and-pencil grasp of math that the ITBS emphasizes. Teachers speak of providing "exposure," or "background" to their pupils, when they describe how they have to teach concepts and skills out of their usual sequence. As one teacher says, "If it's within the first grade curriculum, I might move things along faster in some areas or quickly give them an overview of a concept that would have been taught in May," that is, after the ITBS. This kind of test preparation is usually accomplished with worksheets provided by the district, the textbook publisher, or constructed by the teachers.

Coaching formats. The fourth type of test preparation described by the teachers consists of activities directed not so much at learning or reviewing substantive content, but at instilling test-wiseness in the pupils. Teachers report using district material like MOR, *Scoring High*, materials they construct, and lectures to show pupils what to do in certain testing situations. For example, they coach pupils on how to bubble in their chosen answer option on a machine-scorable answer sheet, how to keep track of their place on the separate test booklet and answer sheet so they do not get off by a line, how not to spend too much time on an item they cannot immediately answer, how to rule out obviously incorrect options and make an intelligent guess on the remaining (although this is by no means common knowledge or accepted practice among the teachers here), and how to use the format in which the correct answer is not given. Teachers often warn pupils that test-makers "try to trick" them with ambiguous words and pictures, so they must watch out. Most teachers stress that, since the test is made of very hard items, that the pupils should not expect to be perfect, but to answer all they can. Alerting pupils to the "godawful names" the ITBS uses in stories and items, one teacher says, "I tell them just to black out the names that are given and put in their own names or the names of their friends, because those names really throw them." Some teachers demonstrate where commas should be placed in friendly letters so that the pupils will be able to respond correctly to those items without even reading the contents of the letters themselves. They provide practice in the format for spelling and punctuation tests that differ from the format used in their texts.

In this type of test preparation, teachers are less interested in teaching substantive content than they are in training pupils to react to the test in prescribed ways to augment their test scores. The teachers attempt to familiarize their pupils with conditions that are essentially foreign to them and to stave off stress and "blocking" that might otherwise result. They talk about using the worksheets "so the kids won't be blown away by the format" or the language employed by the test.

Few teachers believe that test preparation is a blessing. The most common cost the teachers name is time—time taken from the regular curriculum, time from the other activities and pursuits they value. "I give up the flow of the curriculum and go back to review," one laments. "We have to cut back on what our regular program is," "For weeks before, we give up our writing time to do the MOR booklet." In science, "it cuts into their activities time. They don't have as many hands-on things when we're reviewing. They'll have more pencil-and-paper things." Social studies, health, poetry, reading aloud, "the creative side to teaching," "a lot of art-type things," the "fun things that you get the class to work on together," progress in the programs beyond that which the test covers—these are all mentioned by teachers as the opportunity costs of test preparation. Some teachers also note that extensive use of worksheet activities to prepare for the ITBS may backfire because pupils become bored with the repetitive work, and some of the worksheets are so difficult that pupils become anxious and less confident.

> It's just that anxiety that there's all this stuff that's going to be tested and I haven't done it, and my kids are having a great time learning and now I've got to stop. So, you know, you stop and start and stop and start. And I even went so far as to structure my lesson plan so that every 15 minutes I'd hit something else. I'd think, "Boy, I'd better hit something. I haven't hit editing, and it's going to be on the test." Not that I don't think editing is important, but it's the way I would be hitting it. It would have come out more naturally if I had not been stressed out by the test. So I'm structuring the day more. Segmenting it off into those different times of the day where I can make sure that they get certain of those subskills that will be tested. Things have changed. I use more xerox copies. I'll be selecting an objective and task-analyzing it and going over it one step at a time.
>
> The kids sense the change. It was just like we had this wonderful community in here. We all just learned. And then all of a sudden they're sitting straighter in their desks because they can feel this thing from me, this difference. And even with the few map pages that I've been doing that I wanted to see how they are picking up the skills, the whole attitude and aura of the room changed. It's like, "This isn't fun anymore." You know, they didn't realize that they were learning before. But it's just so different. We had this wonderful community in here, and now it's gone.

Teachers' Beliefs about Testing: Interrelationship of Categories

Teachers are sensitive to the discrepancy between true educational attainment and the scores on achievement tests. By the teachers' definitions, testing programs hurt their pupils, and test scores are used more often against teachers than by them. Notwithstanding the perceived low utility and validity of standardized achievement testing, teachers admit to using a variety of means to increase test scores, either at the expense of true achievement, or at least of instructional time that might be spent more productively. Emerging from the data is the link that explains this apparent paradox: Teachers believe test scores will be used to judge and embarrass them or to decrease their autonomy over content and methods of instruction. Therefore they do what needs to be done, and that involves prepping their pupils to take the high stakes test. In their own words:

> Teachers are mastering these tests and they know that perhaps their pay is even based upon the tests. It's only logical that they would do everything they could to increase the scores.
>
> The scores become a reflection on me, on our programs, and our school. So then the reaction is to teach to the test. Which I don't particularly agree with.
>
> I used the worksheets because I wanted my children to do well. And I wanted to make it simpler for them somehow to make sense of the test. And I wanted to keep my literature program. And in the end I probably was teaching to the test. Which is probably what we teachers are compelled to do when pressure is put on us that people say is not put on us.

> And I've had this happen where a child has come in the Friday before the ITBS as a nonreader. Child hasn't been in our school. Child's maybe not been in school, child's maybe been in three or four schools. And those scores go down with my name on them and I am responsible for them. And in that child's records, my name is next to the line that those scores are recorded on. It makes me care less about the results of the tests. Because I'm being held accountable for a set of scores, some of which I'm accountable for and some of them I'm not. And on top of that, it's not my brain that's taking the test. And basically what we're doing is we're assigning responsibility for people's brains. And if Johnny can handle standardized tests really well, and is really smart, and is coming down with the flu, and bombs this test, "Ahhh, Johnny's scores went down. You weren't an effective teacher." And that's becomes a reflection on us, on our program and our school. So the reaction then is to teach to the test. Which I don't approve.

> It's got to [affect teachers' sense of success and self-confidence.] I even get disappointed if the kids in my class do poorly. You know. I wouldn't be able to understand why. [The next year] I think I'd remember the test better. You know. I'd probably try to remember the test better and deal with those things that I thought caused the problems. Like if the majority of the class got low in language, then I'd pay close attention to what's on that test for language. What is it that they're lacking and then the next year, work on that. That's all you can do. So actually you are teaching to the test. Which isn't right.

> I really feel that's where the standardized tests fail us because of the abuses of how we use the results. And the misuse of how the teachers teach to the test. Because the end product may come 12 years down the road...and it will not show up in the standardized tests.

From their personal knowledge and direct experience, teachers recognize discrepancies between achievement test scores and the underlying trait of achievement. Because of the technical characteristics of the tests, the structure of the testing programs, the content of the tests relative to what has been taught, and the characteristics of pupils tested, the indicator deviates from the construct of achievement. They believe that standardized tests reveal only part of important educational attainments. Yet despite this perceived discrepancy and restriction, teachers feel the effects of the test scores on themselves and their students; they strive to raise the indicator to avoid public shame and perceived failure. The means for increasing the average level of the indicator are readily available and heavily used: systematic instruction in the skills of test-taking and content specific to the test. As much as they support literacy and the development of basic learning processes and skills, they view test preparation as a departure from the ways they would normally be spending instructional time in pursuit of valued ends. They have few expectations that activities designed to increase scores will have any lasting effect on pupil attainment, broadly defined. They are encouraged, or force themselves, to direct resources toward activities and goals that they do not respect.

Administrator Beliefs

Do school administrators share teachers' beliefs about testing, or do their jobs or self-interests shape a different view? We employed some principles of theoretical sampling (Glaser, 1978) to shed light on these questions. The aims of theoretical sampling include generating further properties of categories discovered in early stages of a project and building boundaries around assertions by gathering data from different participants or in alternative settings. We found four major dimensions along which administrators' beliefs about testing contrast with those of teachers. These are (a) DEFINING THE EFFECTS OF TESTS ON PUPILS, (b) DEFINING THE EFFECTS OF TESTING ON TEACHERS AND ADMINISTRATORS, (c) ASSERTING THE NEED FOR PREPARING FOR TESTS, and (d) DEFINING THE DISCREPANCY BETWEEN THE INDICATOR AND THE TRAIT OF ACHIEVEMENT.

We include quotations from administrators interviewed in other districts in addition to those in Cactus District, which was the focus of the study. The reason for departing from the original design in this way was to distance the persons from the data they provided and protect their confidentiality. In all, we conducted 15 interviews with administrators at various levels. The quotations we present are characteristic of the categories in the original analysis.

Defining the effects of tests on pupils. Administrators at the central office believe that pupils, except perhaps the very youngest ones, suffer no ill effects because of the ITBS or other achievement tests. To the question whether pupils suffer, "I would say pphhtt!" One states:

> I think it is a bunch of baloney. I am not happy with testing [grades] one and two. I share the same concerns, I think, as the primary teachers do. I think we need to give those little guys the three years to get it together, their developmental rates are so spread out at that point...From that point on, that the frustration and the hyper kids and all that stuff comes from the frustration and anxieties of the teachers. Teachers don't like it, they would rather not do it, it messes up their schedule, and that goes down to the kids. What you say and what your tone of voice is sets up those kids for that hour of testing. And I've seen some set them off...and because of the way the public and the board interpret the results, teachers put pressure on themselves and that translates to the kids.

The beliefs of central administrators contrast with those of principals, who believe with the teachers that the tests are inappropriately difficult. The length and difficulty of the ITBS contribute directly to feelings of failure, frustration, and low self-esteem. One principal says:

> I want the kids to be successful because I want to be able to demonstrate to them you can be successful whatever you want to do. It doesn't make any difference if you are poor, you can succeed. It doesn't make any difference if you have one parent, or no parents, you can be successful and we will help you be successful if you give it a chance. I want to instill that self-pride in them and you can do anything you want to do if you put your mind to it. The thing I don't like about ITBS is that we build them up all year long and then they take the ITBS and because they are not reading at their grade level, they are forced to take this test most of which of them can't take it

> successfully. And so, we basically say to them hey, you're stupid. You know, we build them up all year and then at the time of the test, you are stupid because you can't read the words; you can't analyze the information and you have to guess or you just mark anything you want. We have had a lot of kids do that. That's observable, that's definitely observable, and you see kids crying and upset. They know the pressure is there to do well on this thing, this test, and that they can't read it well enough to be successful.

Another principal adds that the deleterious effects of the test on pupils have to do with the loading of the ITBS with language and background information that are not part of lower class children's experience.

> I think that is the main concern and you know their whole culture and how their families operate, the whole thing, just isn't pictured in those tests. So it doesn't have anything to do with them. Because they are good achievers. I wouldn't be in this school if I didn't believe that. We know they are bright. I think they have done really well this year. I am real excited about what we have done, but that's why I doubt that it will show up on the test. And the kids don't have perseverance when there is failure. It is not one of their best attributes. These kids seem to, they will do anything if they are doing well, but once they get into the mode of "I don't know this," "I can't do this," "I don't understand this," then they fold down. I don't know—middle-class kids do that too, but not quite as often. I think maybe it has to do with on the basis of [their parents' message] to try, try, try again. My kids don't often get that kind of support. I think they are getting it at school, but I'm not sure they are getting it at home. You know, you're stupid if you can't do something, so they have a tendency to stop.

Although administrators agree with teachers that there are too many tests, each administrator seems to have a favorite test that is added to the testing program. District administrators say they have no choice about administering the ITBS and extol the virtues of CUES and BST as means of keeping the teachers attuned to the district scope and sequence. Meanwhile, the principals have added tests that are appropriate for assessing the effects of their local (as distinct from the district) program: The Metropolitan Achievement Test and the Reading Miscue Inventory, respectively, for the two schools in the study.

Defining the effects of testing on teachers and administrators. Central administrators recognize that teachers feel the pressure of the mandated testing programs, yet they believe that teachers have "overreacted" out of proportion to any "real" consequences of low test scores. According to this belief, the pressures school personnel feel in regard to test results are self-imposed and do not emanate from the central office. Another administrator blames the teachers' lack of technical knowledge for the anguish they feel about test scores. If they understood that, to get high ceilings, the first grade math test must include multiplication items, teachers would not be so worried about the results or about exposing first graders to multiplication.

About the overreaction of teachers and principals, one district administrator says:

> I guess in the nine years I've been here I've seen the perception on the part of teachers gives testing more importance probably than is placed on it at the district level, and I can understand that. You know, the board looks at testing and makes statements about it and the district's being high or low or whatever. The state puts out a booklet with everybody's scores in it so parents get hold of it and says [an adjacent, affluent district] is higher, which puts a pressure on teachers because they want their kids to do well. [District administrators] are not asking teachers to push kids. [But what] we've seen happen this year is teachers teaching inappropriate skills in inappropriate grade levels because they know there is a question like that on the test.

Although district administrators lay the blame for testing effects elsewhere, when pressed they acknowledge that there may be grounds for teachers' and principals' fears. For example, standardized test results are one criterion in the evaluation and merit pay decisions of principals. The district administrators minimize the importance of this, saying it is "just one of many factors," but some principals fret. About principals' reactions, a district administrator admits:

> [Principals] are accountable for getting some results. We have structured a situation, and we want students to show growth on the testing, we want students to show growth through the curriculum. We have structured that evaluation focus so that it looks at two things: first of all, it looks at growth of students so that it is relative to where kids start the year. We look at patterns of growth within a school. And secondly, it is comparing the pattern of growth from school to school so that the patterns are similar. And a way a principal can really be penalized in any way on the evaluation system is if there was a significant decline in the pattern of growth within a school. We look at grade level to grade level comparisons and we are capable of disaggregating the statistics so that background factors are considered. In our school profile we look at students who are continuous in the school during that evaluation cycle. So it is pretty fair. You know, we assume that if kids are well instructed that they will show growth that generally approximates the growth being shown in other schools in the school district. If something happens, you know, that says kids across the district are showing month to month growth, but in this school it is only six or seven months, and we disaggregate that data and find that some of the excuses for that don't really hold water, then we would take a serious look at what is going on in the school.

Another says:

> I've watched the principals in the past. Those who try to relax about it and not put too much stress on it, and they work with their staff, and then it turns out well. I have seen others where you know, the end of world will come if we don't reach such and such a level on the test. But look at the state focus. You know what is going to be in the newspaper coming in May. You can see it for days. So if you are a teacher out there, just trying to relax

> and do your best, then you start seeing these scores [in the newspaper] slap you in the face and you would react to it.

A district administrator claims that the district's only interest in the test results is that the personnel in the schools should make use of the information for curricular and management decision-making. Yet, this person acknowledges that teachers and principals, "Look at test scores, and my test scores better be higher or the school board is going to be on my back. And they have seen people disappear because board directions were not followed or test scores didn't look all that great. Specifically in our language arts area."

Principals align with teachers in their beliefs about the effects of testing, and feel the added tension that, if their schools score low, the district will hold them responsible. They recognize that when scores are low, "it's a reflection on you." One says:

> Those numbers seem to be magic. You know, we live and die by those numbers. There's a lot of attention paid to those numbers. If your numbers don't come in right, then that is one of the main measures of your effectiveness as a principal. When the ITBS scores come out, there are administrative team meetings as soon as the data is collected and summarized. They will get feedback on each building and they will have printouts that compare school by school: absolute scores school by school, gain scores, a variety of kinds of analysis that they can do, and you get ranked. Just completely rank them the highest scores at the top and the lowest scores at the bottom and you look at the list and you can see where you've scored. At the bottom of the list you have our school and a few others, and schools that have the highest percentage of low income students will come in at the bottom. And then you'll have the extremely high income schools at the top, and they will be patting themselves on the back because their students have done so well. The other schools at the bottom will be saying, well our schools are bad because the scores are low. A lot of people recognize that there are probably economic factors that have a strong effect on the outcome of these measures, but it doesn't seem to change the fact that it is still viewed as one of the main—they say there are several variables—but one of the main variables on your evaluation of the school's administration. Then they will look at gain scores and when we look at gain scores, then you hope they provide a different view of what the school is doing. If the students have moved close to a year at each grade level no matter where they started, at least they are progressing a year's growth with a year's span in school. And that is good feedback. [Because the scores are] treated so seriously, you have to take them seriously. Either that or you go someplace else and play by some other rules.

Another notes that the competition among schools that comes from ranking them on their ITBS scores is less severe now than it was in the past, and the Cactus central administration attaches less significance to them than do some other Arizona districts.

> Under the former regime, there was more competition and there was a little bit more pressure put on principals to "perform or else" kind of thing. And I think what the district is trying to do is be a little bit more subtle about it and

> be a little bit more sensible about it in terms of letting principals take the rein and make some impact at the local school level and focus in on that rather than worrying about where everybody else is compared to them. So I think the competition aspect has changed quite a bit recently. And now the competitive drive comes more from within the person.

Although the district has soft-pedaled the test-score ranking this year, these data are public and accessible to the media. The State Department of Education publishes ITBS data by school and grade (Bishop, 1988), so that newspaper editors and others may readily rank schools by scores. Therefore, concern about invidious comparisons of schools based on ITBS scores is more than illusions, despite what district administrators claim is their policy about use of results. Principals believe that the obsession with scores starts at the top and works its way through the organization. One says:

> Well, I think it comes from the legislature. I think you have legislative pressure which apparently comes from community pressure that there is the perception that schools are not effective, and we can look at the history for the last 20 years and it may very well be that that has been the case in some places. But, you know, who really looks at that data? People probably who are removed two or three times by the schools probably place more emphasis on that information than people who are closer to the systems. If I was totally removed from public education and I wanted to gain information about the schools, in general, in Arizona, yeah, I might go out and visit schools but I couldn't visit them all, but I could look at the data and make some decisions. And it is politics, too.

Another adds:

> I think the district has very high standards for itself. It is an ambitious district. They have a tradition of seeking educational innovations of various kinds and are always striving to get there. They are very concerned about public perceptions about its excellence. You know, even the real estate people have this book that ranks the schools on ITBS scores. So this is more pressure on us.

The push for public displays of accomplishment at the district level puts historically low-scoring schools at particular disadvantage:

> If I was in a higher performing school, I might be singing a little different tune. Okay? And if my kids were really super achievers and if we were way, way above state and national norms, I might really say, gosh this is great, but we are at the other end of the ball game, so I'm more skeptical of its relevancy. But I think the pressure comes from the top down. The teachers, you can see the pressure mounting two weeks or so before, even a month before ITBS, "we got to practice, we have to review those skills." Well, we take the test and a week after ITBS, "whew it's over!" And the very nature of the way teachers react you know that there is no real strong sense that this is meaningful other than to see when the test data comes back how we did.

Another principal confirms that the emphasis placed on scoring high on tests is a fact and not a paranoid delusion:

> I don't want to make it sound like we are not affected by it, we certainly are, and yeah, I think every principal in this district, probably every principal in this state, wants their kids to do well and works hard with their teachers to prepare them to take the ITBS. We all want them to do well for ourselves as well as for the kids. Because, you know, there are some things that hang in the balance, I mean, let's face it. If my kids really did very, very poorly on ITBS after doing, compared to our standards, fairly well last year, people would ask what is going on. What is wrong? You know, over a period of time, if the data kept coming up really poor, you look at that and no one is going to brush that aside. It is going to have an impact somewhere along the line. Certainly the district is going to look at it, certainly I am going to deal with it, because it is at the very least, an indicator that something has happened good or bad."

The theme of diminished autonomy that can result from low test scores echoes in this passage:

> I think that one of my greatest fears is that after three years they will look at this and say, "we have given you three years to show progress, and where is it?" That is a fear of mine. The only thing I know to do is what I am doing and that is working as hard as I can to show by other assessments that we are being successful with kids, because we can't count on how the test scores are going to come out. I have to seek other ways of making our school visible and making what we do credible.

Asserting the need to prepare for tests. Teachers believe that special preparation for the ITBS is necessary. Not to prepare is almost a deviation from the norm. Yet central district administrators disagree sharply, claiming that teachers should devote only very brief times to test preparation, solely to practice in darkening answer sheet bubbles and following test directions. One speaks for all:

> You see, I think, especially in the math, if they have been using our math program with the systematic review which is a built-in way for them to review all the skills that have been previously taught—you know it is just an accumulative kind of thing—they don't have to do any preparation for the test. If they are taking their kids from where they are to where they can go, and they are doing it in that systematic way, with that review so that the past skills are maintained, then all you have to do is give the test in April. You don't have to do anything special. I don't think we have to do any more in reading; I think it needs to be an ongoing preparation in all of it in terms of test-taking skills. I think we need to do some time tests with them throughout the year at different times, to get them used to that. Other than that, I don't see how you prepare for it. The objective list that they let us see for ITBS, it's fairly general, you really can't, you can sort of tell, you know, it is not a specific content-laden kind of thing that you can really work it. They have a correlation of what our CUES, our Scope and Sequence, what correlates to make sure you teach those objectives. I don't think they need any—if you just teach the scope and sequence, they would be fine. And if you've got a group of kids that didn't make it through the scope and

> sequence by April 11th, they didn't make it. You did the best you could. They keep saying you give the ITBS too early, all education stops. Why? Is ITBS the end all? But for some of them, its all down hill from there. And yet, we have a good six weeks of curriculum to complete. I don't think they need any special preparation for it.

Principals are obviously in the middle, and their stated beliefs reflect their ambivalent position between the teachers' almost feverish need to prepare and the district administrators' official stance that special preparation for the ITBS is unnecessary. One principal states that the decision to devote time in this way is left up to the teachers. In the following statement, the principal seems to underestimate the extent to which teachers alter their instruction to get ready for the tests, and wants to stay in the dark:

> I actually wanted to use the *Scoring High* format for practice and it goes back to what I said earlier. These kids are not real test smart and need to be taught how to take tests. We did not actively go out and do that. Teachers did use some of those MOR materials [distributed by the district]. They did review a lot of the concepts that they had taught during the year. We didn't do a lot of formal test taking practice, which probably we should have done. I bet we could just by teaching them a little bit more effectively to take tests, fill out the forms, and make sure that keeping everything in sequence, it might have helped. We concentrated on the skills we taught and I didn't put a lot of stress on teachers in terms of how they should do it, or what they should do. I pretty much left it up to them. We cover a lot of that sort of thing in the Study Skills book. We didn't have a real formal school-wide focus on practicing for the ITBS. I pretty much left it open to them at the time. Because by the time the test comes everybody is so tightly wound that it goes completely overboard.

Defining the discrepancies between the indicator and the trait of achievement. We questioned administrators, just as we did the teachers, about what the mandated tests reveal about pupils, programs, and teachers. The differences between the teachers' beliefs and those of central district administrators generated several original properties of this category. Teachers sense acutely the discrepancies between the scores attained on ITBS, BST, and CUES and the trait of achievement as they define it. Teachers dwell on the distortions and fallacies of the scores and produce elaborate anecdotes to support their claims. They often become animated and emotional about how testing programs falsely shape public perception of the qualities of public schools. Yet this issue draws only a passing mention by central administrators. One of the latter refers to the ITBS, for example, as a "gross comparison of where we are in relation to the nation," without specifying the trait on which the district is being compared. Another calls the ITBS a "dipstick of how we compete with the nation," and the BST, a "dipstick of where we are in our curriculum." Their comments seem to assume that it is not the "what" that matters, but only the "how much?" Another administrator casually calls the test scores, "ballpark figures" and refers to the group gain statistic as a measure of "growth" the schools forge compared with the average progression of grade equivalent scores published in the ITBS normative data. For the central administrators, what is growing or what is being "dipped" into is whatever the ITBS or BST measures.

These contrasts between administrator and teacher beliefs and the differences in amount of discourse devoted to this issue led us to conclude that, compared to teachers, district administrators gloss over the discrepancies between the trait and indicators of achievement. Teachers believe that achievement test scores ought to carry information about real achievement. Differences among scores in a particular classroom ought to reflect the real differences among pupils. For district administrators, the quality of the information is not a burning issue. Yet administrators often refer to the scores themselves, as one says, "turning them inside out and examining them from every possible angle." What they look for in their analyses are any patterns that might reveal absolute or relative declines, differences among schools, grade levels, subtests, and differences between the group gains that are made and those that the central office asked the principals to estimate. They know that if they do not examine all these declines and differences the board members and media will do it for them, probably in an embarrassing public forum.

Whether these numerical changes and variations are meaningful is a question that is largely overlooked. One testing specialist reports that in the past he has tried to inject technical issues into the analysis, but to no avail.

> When I presented results to the board, or to principals here, and most of mine have been to administrators within the district, someone on the board will immediately jump on the numbers and ask how come [the neighboring district] got 3.0 and we're only 2.9 in reading at the second grade level. I look at him and say (I don't say shit, I do clean up my language a little bit when I speak to the board), I say, "It is not significant, just because they are .1 above us, you know you are talking about 200 kids at a grade level, and we are talking about 2,000 kids, and we are looking at a .1 difference. It is not significant." "Well, it cost me a dinner," this is what one of the board members said, it cost him a dinner because our district was lower. But people, not just the board members but others here—they look for those differences. And sure. They will find it. If the scores are equal, they will find it. Out there, a .1 difference is almost a catastrophe.
>
> You also have to look at the gains with a certain amount of skepticism. I have said to them, when you look at gains like this, a .2 below your expected gain of 1.0 isn't a significant difference, so don't be looking to remove teachers or change programs with a difference that small. Only look at those where it is greater than .2. On the other hand, don't be elated because it is a 1.2 as compared to a 1.0. You have to keep your statistical limitations in your head. So I do know that lay people, board people, legislators, and what have you all are talking about excellence in education. What does the district do when our gains are above 1.0, where do you set your goals? You set impossible goals, you know, we now say to a school, okay now we not only want one year of gain, we want a 1.2 year gain. You know, where does it stop? There has to be some point. The district can only go so far up and you know, there has to be a leveling off and you just can't constantly be changing your gains. They've got their five-year strategy planning or something, and a couple of their goals as far as I am concerned are unrealistic goals. You can't reach them.

Hearing but not heeding this solitary voice, central administrators overlook or choose to ignore the advice about technical issues of testing: unreliability of gain scores, ceilings on the amount of gain possible, the insignificance of differences between subtests, schools, and districts, and the unreliability of the tests (especially the district CUES and BST) themselves. In spite of these technical problems, central administrators encourage principals and teachers to raise scores that are low and promise the board and the public that schools will exceed earlier gains.

Administrators use the district testing program as organizational tools: a way to make sure all schools adhere to the District Scope and Sequence. One central administrator ponders how the district's criterion-referenced testing program should function:

> I look at some of the reports we generate for schools in terms of CUES and you can't understand the relevance of some of the numbers on the page and that causes a great deal of trouble. If I can't see what that number means or if it is open to interpretation, or it is not totally clear, then our tool we are producing for principals to manage schools by and teachers to manage classrooms by is compromised. And that doesn't justify the effort we put into assessment. We have to ask ourselves how the testing system serves teachers as well as administrative functions, because that is what gives it value. And very often, the central administration has been the sole determination of when a test was given, what was in the test, what would be tested, what areas would have tests developed for them. And I think teachers are capable of helping us with some of those decisions, I really do. Now, we have some things that we have to do. My job is to supervise the principal and the principal really needs to know how the program is being managed in the classroom. That principal needs a management handle. I think CUES can present that information, but the same set of criteria for CUES needs to apply as I think needs to apply in basic skills. The approach this year has been, OK, let's talk about testing, so we start talking to principals and they said the best thing we can do for it is do less of it. So we backed off, for example, in CUES and we didn't require a report date until the end of the first semester. And the results of that was the first formative curriculum report based on our testing system that the principal had in his hands came along late February by the time the Data Processing Division got it out. Is that early enough? I have some real questions with that. How are the people that have to deliver the instruction that are responsible for managing the school site system, is that information supporting their need to know as well as guiding their decision-making in a way that they need to make decisions about kids that need to be made.

A central administrator describes the role of the district-mandated tests as the final standard for judging a school's accomplishment of the district curriculum. A school staff might decide to pursue a program different from the one prescribed by the district (e.g., in math). Yet the district has determined that each grade must accomplish certain skills in a certain order. The Basic Skills Test scores would show that the school that taught math in a different, but equally effective way was deficient in math. The central office would take the low scores as a sign that the school ought to bring its math program into compliance with the common model. There is building autonomy, the central administrators claim, yet only in terms of the materials and methods by which the common content is taught. The substantive objectives of the district, as reflected in the CUES and BST, are not subject to negotiations or variation from school to school. This is true in spite of, as administrators admit, the inaccuracies and technical inadequacies of the district tests and statistical reporting of results.

> I had some teachers who just didn't believe in the CUES test. They didn't think they were valid, all that stuff. And I said fine. Then you validate for me that the kids know those skills. Make your own test. If you find a test in the book that you think is better, I said, fine, but I need to know from you that those kids have mastered those skills. I said, now you know, my other measure of that is when we give the Basic Skills Test those kids should come out with mastery. And I think that is something that you need to look at with the CUES. You have to use the test we prepare. We prepare a test to make it easy for you. If you don't have faith in that test and you have a better way, go ahead and test it, but remember, you are accountable for saying the kid has that skill, and then when we Basic Skills Tests at the end of the year he should be able to perform. It has to be measurable.

Pupil Beliefs

What do school children believe about tests and test scores? We examined this question in several ways: by recording their comments about tests during our observations of their classrooms, obtaining secondhand information from their teachers, conducting and recording group interviews with a fifth grade class in one focal school and a sixth grade class in the other, and having primary teachers solicit journal entries on the subject from their pupils.

The results are consistent. To show the beliefs of the fifth and sixth grade pupils, we reconstructed the following conversation from their statements.

> How do you feel about these tests?
>
> They're all right. They're not the best tests. You do all this work [in class] and then when you get to the test, you forget it because you have so much work and they don't give you enough time. They should give you as much time as you needed on the test instead of just timing you. Then you wouldn't have to worry about rushing through it and you can think about the problem you're doing. It would be a lot easier and you could concentrate better.
>
> They give you a test to learn and then don't give you long enough to do it. And you have all of these pages, like 50 pages you have to do, and all of these questions, and then you have to try and rush through it and you get everything and you don't do your best. You're rushed, and you forget what you're doing. You do, like, whatever comes into your head.
>
> What happens then? What difference does it make if you don't do your best?
>
> The difference is that the Iowa test is supposed to tell you what you know. So if you rush through and forget everything you know then they will think you need to learn everything over again, so they teach you again even though you really know it.

> Why would anybody get nervous when they take a test?
>
> Because it's a test. And when you take a test you get nervous because sometimes you don't know the problem and you just get nervous when you have to deal with the problem. When you go through you have nerves because it's a test and you want to do well.
>
> Why would anybody want to do well?
>
> So you can pass.
>
> Because at the end of a year to do the Iowa test, if you get them all wrong, it shows them if you should pass or not, and they want to do well on the test to pass and to get to the next grade.
>
> Do the tests show anything about you?
>
> How we think.
>
> It tells you how we're doing.
>
> From the Skills, they see how good you are.
>
> It shows that you've paid attention. It shows how you react to certain things, too. I like math. And I don't like certain subjects. So I'm good at math and I'm not that good in social studies. And the test shows that.
>
> It would tell you how good the teacher is. If they make it interesting the child usually likes the subjects, so they'll try harder in it. But if the teacher makes it boring or hard work or, you know, the kid doesn't want to do it. So, obviously, they'll get a lower grade in it than they would in a subject that they like.
>
> It tells us how hard the teacher teaches.
>
> What it tells about the teacher is if the teacher has been teaching the kids what they're supposed to know. And how the teachers have been teaching, like, stuff so that the kid—what grade the kids get depends on how good their teacher was teaching them. If a child does bad on his subjects the teacher is going to give him more homework, which means he'll do more studying, so he can get a higher test score.

Except for the time constraints of the tests, pupils believe achievement tests reveal a straightforward picture of their year's acquisition of skills, mental competence, subject matter knowledge, and their teachers' performance. They believe that test scores will determine whether they are promoted or retained, and low scores will precipitate teaching of material already covered and more rigorous coverage, including more homework. Despite some report of nerves, anxiety about performing well, and frustration about the tight time limits of the test, they seem to believe that testing is a normal part of the business of school, they adjust to it, and they report no ill effects over and above the nerves they feel before any kind of test.

In about equal numbers, the primary pupils write about the fun or the boredom of testing, its difficulty or ease, their belief that they did well or poorly on recent tests, and the wasteful or fruitful use of their time to take tests. At least a tenth of the pupils writing journal entries mention that the tests make them feel stupid. Compared to their teachers, the primary pupils paint a slightly rosier picture of the effects of testing. We reproduce a cross-section of their journal entries on the next few pages.

Public Beliefs

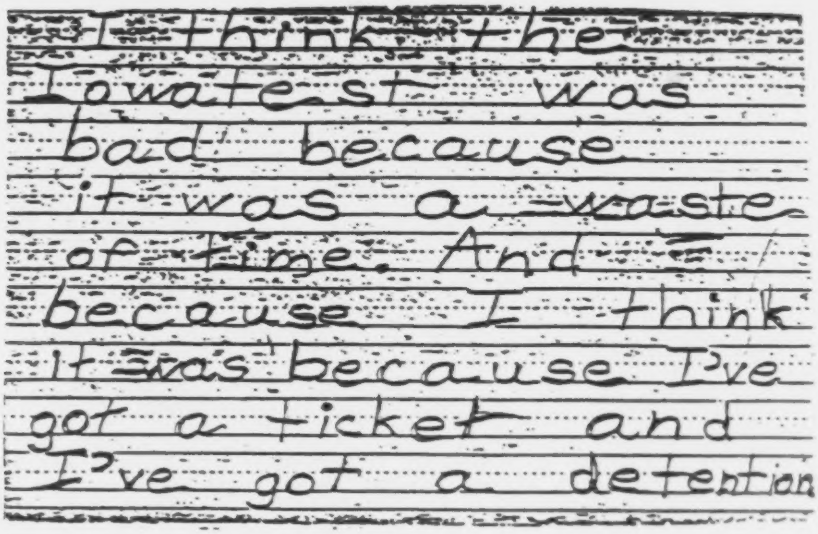
To understand educators' beliefs about testing, it is necessary to scan the social environment for supporting or contrasting statements made in the public forum by policy-makers, newspaper editors, members of boards of education, and laymen. While this study was in progress, debate over proposed legislation to eliminate mandatory testing of first and twelfth graders offered a forum for airing the public's ideologies about testing. Although the legislation eventually passed, one can view its success as either a major shift or temporary aberration in the public sentiment in Arizona about schools and the role of school testing. We characterized that view from these different sources of data and reconstructed it in the form of an interview with an education editor of a major newspaper. An actual editorial that the reader can compare with the reconstruction is reproduced in Appendix B.

> What is your view of the status of Arizona schools?
>
> There can be no doubt that educational quality is universally dismal in Arizona schools and is steadily deteriorating. Test scores show this. Teachers, like other public employees are inept and possibly lazy. As leftovers of the liberal sixties, the curriculum is full of fluff and less academically oriented than in it used to be. The "educational establishment," which includes teachers' associations, the State Department of Education and its chief officer, State Board of Education and all three colleges of education, is inept, self-interested, and possibly corrupt. To give you an example, the legislature tried to modify the rules for getting a superintendent's certificate. They specified that the candidate could come from a business background but had to take a number of courses in educational administration courses at our universities to keep his certificate. We likened that to Chinese water torture for someone of the CEO type. Naturally, the establishment came out to try to protect its own interest. For another example, the establishment always supports greater public funds for schools despite the evidence that test scores have plummeted during the same time that budgets have skyrocketed. No matter how deep the crisis, money will not solve the problems of schools. Research proves this. The correctives for bad schools are sound business practices, high expectations for pupils, competition between schools and districts based on parental choice of schools, and an emphasis on teaching of phonics and moral character. But first you have to have the information, and that's why the tests are important.

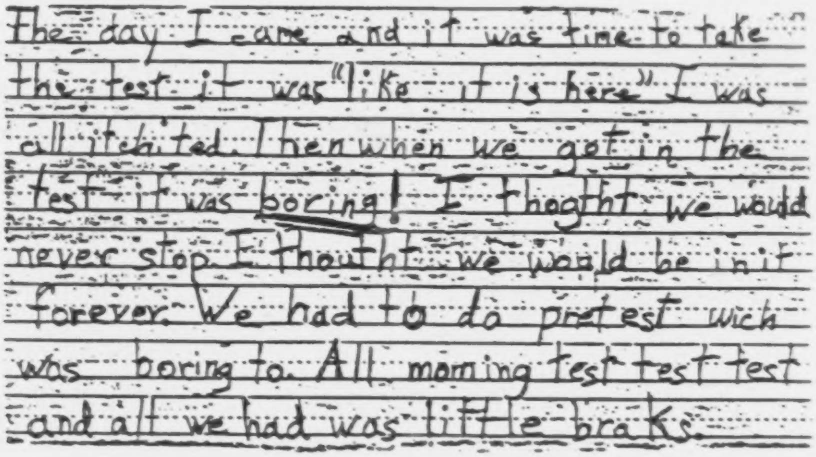
I feel like thet ITBS test wast
a very fun and that I hope
that I do it a gen I
now that I will do it a gen
I hope that I go to
threed grad but the ITBS
test was fun.

I was right, the test stinks.
It aught to be burned.
I should go and play soccer
with it.

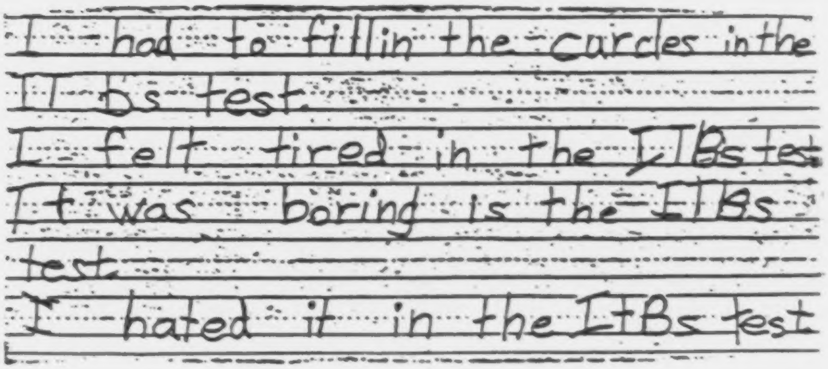
April 11
I hate The Test
I am a frad
I will fall in
frist agen.



I think the
Iowatest was
bad because
it was a waste
of time. And
because I think
It was because I've
got a ticket and
I've got a detention

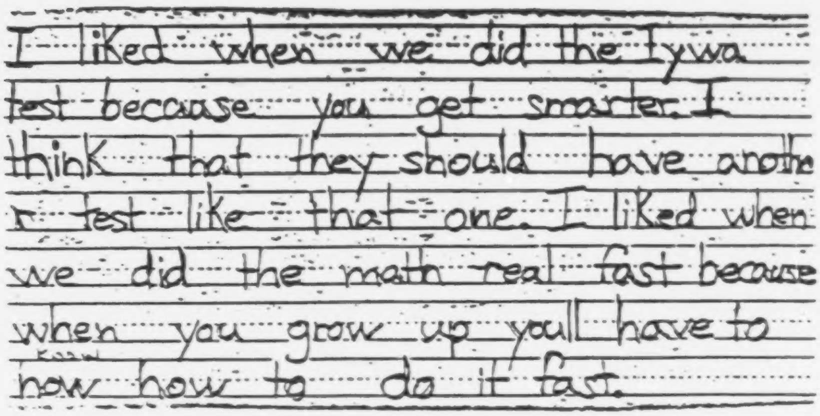


The day I came and it was time to take
the test it was "like it is here" I was
all itchited. Then when we got in the
test it was boring! I thoght we would
never stop I thouht we would be in it
forever. We had to do pretest wich
was boring to. All morning test test test
and all we had was little braks.

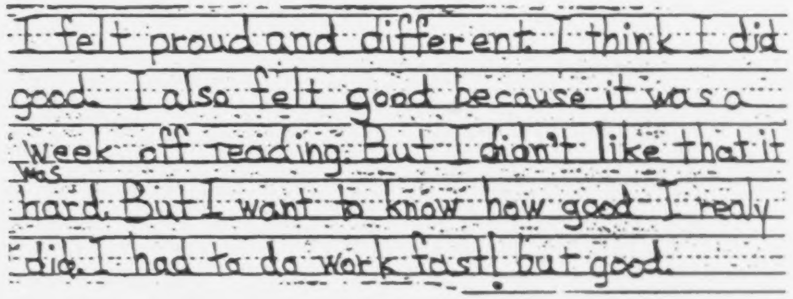


I had to fill in the curcles in the
IT Bs test.
I felt tired in the ITBs test.
It was boring is the ITBs
test.
I hated it in the ItBs test

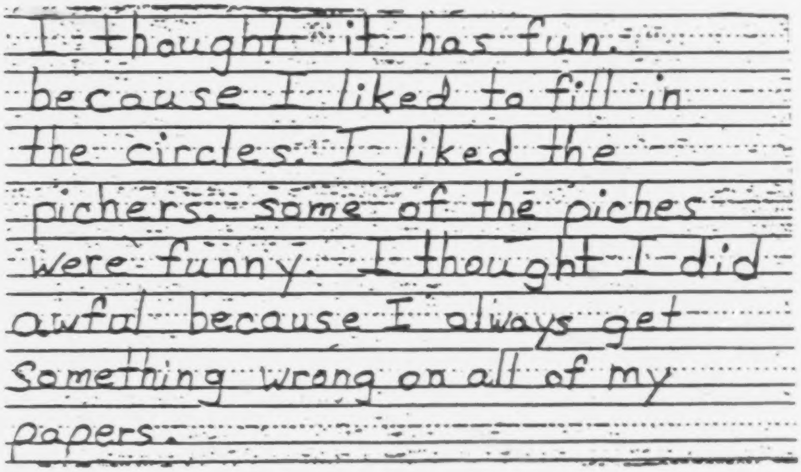
BEST COPY AVAILABLE



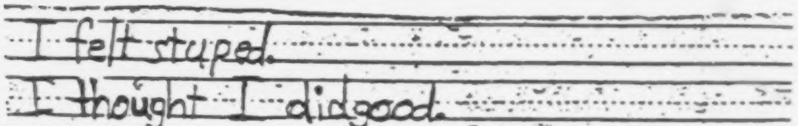
I liked when we did the Iywa
test because you get smarter. I
think that they should have anothe
r test like that one. I liked when
we did the math real fast because
when you grow up youll have to
know
now how to do it fast.



I felt proud and different. I think I did
good. I also felt good because it was a
week off reading. But I didn't like that it
was
hard. But I want to know how good I realy
did. I had to do work fast! but good.



I thought it has fun.
because I liked to fill in
the circles. I liked the
pichers. some of the piches
were funny. I thought I did
awful because I always get
Something wrong on all of my
papers.



I felt stuped.
I thought I didgood.
And I thoughtI was stuped.

What is the role of testing?

The role of tests in this picture is clear. We must test every child every year so we can know how Arizona pupils compete with children across the country. If you analyze them correctly the Iowa can measure the yearly progress of individual students and provide comparisons of the effectiveness of schools, programs, and policies. They form the basis of a rational system of teacher assessment based on student outcomes. Recently the State Department proposed that we do away with the standardized test and replace it with tests constructed by the Department based on essential skills. They would be administered only to samples of students. To me, this is like turning the hen house over to the fox. They just want to make Arizona schools look good, better than they are. If you think that the state board will produce challenging objectives and tests, you just don't know the political history of education in this state. The only true comparison is on nationally normed and standardized tests with pupils throughout the country. Proponents of that testing bill also say that it will save money. To me, our testing program is the best bargain we have. They also want to do away with the Iowa because it hurts the tender psyches of the children. Bring out the violins. Doing away with the test would deprive parents and educators with information they have a right to know, yearly progress of their children and their children's schools.

Why are test scores important?

Test scores correspond to economic competition, research shows that. On international comparisons we rank right along with third world countries, and we look extremely bad against our greatest economic competitor, the Japanese.

There is a lot of emphasis here on reporting test scores by schools. What effect does this have?

Public publication of low test results will expose pockets of particular ineptness. You might call it applied anxiety. So if the teachers feel pressured by the test scores, all the better, because maybe they will apply themselves more diligently. If the results of standardized tests are not made public, teachers will work even less hard and will teach content that is even less essential than what is now taught. Publication of results also supports an ethic of choice. Under such a system, parents will avoid schools where low scores show that the schools have failed.

What do you believe the Iowa measures, and what about the issues of ethnic bias that teachers raise? Aren't there other indicators of successful schooling?

The test tests essential skills in straightforward ways. Raising contingencies such as test bias for minority children is just a red herring the establishment uses to justify the failure of schools. A prominent superintendent recently quit his position, and we wrote that it was time for a new focus in that district. Although he raised money for the district, successfully implemented a desegregation order, began a promising magnet program, and the drop-out rate decreased, his achievement test scores did not go up. We need someone

> like John Murphy of Baltimore who will emphasize achievement. He set a goal that was more than establishment jargon, that the standardized test scores will be above the 75th percentile by 1990 and the gap between white and minority tests scores will be reduced. That is setting high expectations for achievement. The gains he already produced are the largest in the state, even though the drop out rate and SAT scores haven't been affected.
>
> What will it mean if the bill to eliminate first grade testing does pass?
>
> It would mean that teachers, the union, and education professors used undue influence in lobbying the legislators, who bought their fabricated stories of harm to children. The real story is that they don't want to look bad, the test makes them look bad, and so they are on a concerted campaign to eliminate standardized testing, starting with the first grade test.

The beliefs laymen hold about testing are no doubt less extreme than those reconstructed above and may be peculiar to Arizona, a conservative state. But public opinion polls taken in Arizona as elsewhere indicate favorable attitudes toward testing: Citizens endorse the use of standardized tests to certify teachers and judge the effectiveness of schools and teachers. The State Department publishes the ITBS scores of all schools, a document that is widely requested and read. Relocation services and realtors use the test score data (*Metropolitan Profiles*, 1988) to help newcomers select a neighborhood in which to purchase a home or rent an apartment. The commentary of the editorial writers is rarely challenged, except occasionally by members of the professional education "establishment," who are apt to draw a sarcastic rejoinder.

The majority view holds that the test score adequately represents the accomplishments of pupils and, in the aggregate, of teachers and schools. An improvement in scores is almost universally viewed as an improvement in real achievement, and the difference in average scores between two schools reflects real differences in their relative effectiveness. Even if teachers or schools are hurt as a result, the publication of test scores is a matter of public interest, according to this view.

Beliefs of Testing Professionals

One might imagine that testing professionals—that is, psychometricians—possess knowledge in the ideal sense rather than beliefs about testing. After all, they command reliable, consensual knowledge gleaned from psychometric investigation and theoretical analysis. For example, on tests of achievement such as the ITBS, the average scores of Anglo students are higher than those of minorities. This knowledge is well-substantiated and not in dispute among testing professionals. What one may infer from this test score difference, as opposed to the difference itself, is a matter of contention. Can one infer that the difference is due to bias in the test itself or, instead, does it accurately reflect genuine differences in the educational accomplishments of the two groups? Alternative interpretations such as these are matters of belief. Should society judge a school's performance in light of the background and ethnic groups the school serves? Or do we do a disservice to minority schools by failing to hold them to a uniform standard of achievement, that is, one that is "color blind?" These are matters of value and group interest, based on inference from the data. Therefore, we represent as beliefs the writings of testing professionals and contrast them with beliefs of educators.

Upmost among the beliefs and values of testing professionals is the responsibility to promulgate the professional standards established for mental testing. That is, most professionals trained in psychometrics strive to oversee testing practice and make sure it follows the principles and standards agreed upon by the three professional associations: American Psychological Association, American Educational Research Association, and the National Council on Measurement in Education (1985). Although these standards cover such things as ethics of testing and qualifications of test administrators, the standards most relevant to the present study have to do with the validity of tests. The most salient belief among psychometricians is that evidence must exist that a measure adequately represents the construct (Cronbach, 1971; Messick, 1988), in the specific context of its use. Applied here, this means that researchers must establish a close relationship between the indicator and the trait of achievement. It means that users of the ITBS must show that the test adequately represents the construct of achievement in comparing the effectiveness of schools, placing pupils in grades, tracks, or programs, communicating to parents about their child's yearly progress, and evaluating teachers. For testing professionals, background factors such as pupils' sex or race that correlate with achievement test scores detract from the construct validity of the tests in all the uses mentioned. Pupils' differential anxiety, fatigue, and level of effort also work against the validity of the test, which assumes constant amounts of these states among all tested groups (Haladyna, Haas, & Nolen, 1989).

Discrepancies between the content of the test and the content of the local curricula also imperil the inference that nationally standardized, norm-referenced achievement tests adequately measure pupils' attainments. This issue particularly galvanizes advocates of criterion-referenced assessments. They believe that tests like the ITBS are so general that they can test only a portion of a school's curriculum. They believe that the need for high ceilings and high item discrimination levels (wherein the ideal items are those that only half the testing population gets correct) makes it necessary for tests like the ITBS to be too long and difficult, thus creating feelings of frustration and failure among children. Those professionals who advocate norm-referenced tests dispute these claims and suggest that benefits of such testing outweigh the drawbacks named by their professional rivals.

Finally, testing professionals believe that districts' or teachers' use of special programs to prepare children for the ITBS is "unethical" and "illegitimate" (Haladyna et al., 1989; Mehrens & Kaminski, 1988). Coaching pupils on specific test items, reviewing curriculum known to be covered on the tests, or using programs such as *Scoring High* undermines the relationship between the trait and indicator of achievement. In a class that has prepared especially for the ITBS, the test score means something different than a score from the same test in a class that did not prepare. Furthermore, testing professionals would say, since some schools use these techniques and others do not, one can not validly compare them. One could not infer that the difference in average scores attained by the two schools was caused by the preparation activities or the real quality of their teaching, curriculum, or the real accomplishments of their pupils.

Among advocates of criterion-referenced tests (CRT), test preparation is not considered unethical since no inference is made to a universe of achievement, as is the case with norm-referenced tests. Advocates of CRT assume that their tests address skills determined essential by school, district, and state, and therefore practice for the test is tantamount to legitimate practice of the skill itself (Cohen, 1987).

The relationship between the testing professional and the educational organization that administers tests and report scores is distant, but one can discover a few links. For example, finding its testing program failing or criticized, a district or state agency might call in a testing professional to make a study and some recommendations. Professionals aware of abuses in testing practices or incorrect interpretations of test results might write letters to newspapers or boards of education, bring cases before boards of professional standards or ethics, or testify in court cases. Papers on the topic may appear in practitioner journals and eventually filter down to the administrators of districts. Educators may remember a little of the principles of testing from their college days. Some districts hire testing professionals in staff positions.

In all these ways, the beliefs of testing professionals can play a part in deliberations about testing programs. Yet as our findings show, their message has little chance of directly affecting testing practices. Teachers and administrators do not belong to the same organizations and do not honor the same standards. Even if they had much psychometric knowledge, there is little to suggest that school officials would behave rationally and incorporate the same values, especially if these values conflict with what school officials perceive as the realities of local politics, management of the system to meet common goals, and maintenance of status. They are more likely to view the need to increase test scores as more important to keep their school organization intact than the rather nebulous goal of maintaining the integrity of the inference from the score to the trait of achievement.

Beliefs of Test Critics

Organized critics of tests such as the ITBS also exist outside the school organization. Their beliefs are brought to bear only indirectly, through newspaper articles, op-ed columns and editorials, testimony to state boards, legislatures, and the courts, and political action. We identified and sampled the writings of three groups: the professional lobby, the FAIRTEST group, and the Cannell group.

The professional lobby. Through its national spokesmen, teachers make public pronouncements of their beliefs about testing. In his column in *The New York Times*, Albert Shanker (1988), President of the American Federation of Teachers, writes why he no longer supports standardized testing.

> Since the reputation of a school, its principal, its teachers and the school board and superintendent depends largely on these test scores, schools are devoting less time to reading real books, writing essays, and discussing current events and more and more time teaching kids strategies for filling in blanks and choosing the answers to multiple-choice questions. This destroys much of the value of these tests, which only tell you something if they are an independent measure of what the student knows. The usual test for blood pressure is good only when the patient has not taken medication designed to lower his blood pressure just before the test.

> School districts are now engaged in a process called "curriculum alignment." That means that course content, textbooks, lesson plans, etc. are all being geared to items that will be on the test. These tests only tap a sample of skills, so it's possible for kids to do well and still not be able to understand real books. But since there's only so much time, schools now minimize or totally leave out those things that are not on these tests. This is the tail wagging the dog. Schools and teachers should not be pressured to drop content they believe to be valuable just because it won't be on the test.

Shanker also cites issues of excessive costs of testing programs in relation to information gained from them and usefulness to teachers and others, excessive time devoted to test and test preparation, and psychometric concerns such as the outdated norms of tests like the ITBS, and the need to develop more meaningful procedures for adequate public disclosure about tests and testing practice. In his public statements, he closes ranks with the National Education Association, which some years ago publicly demanded a ban on all standardized testing.

In Arizona, where union membership is small and the professional associations are neither strong nor vocal, the organized resistance among teachers against tests takes place in a grass-roots group called Community for Effective Student Evaluation (CESE). Teachers formed this organization in 1987 to study existing testing practices, increase awareness of existing and alternative testing programs, and modify state laws through political organization and dissemination of information. The organization's beliefs about testing, stated in its brochure, are the following: test publishers decide what is important to test, define what achievement is, and dictate what schools will teach; tests such as the ITBS are harmful to young pupils; test scores are used against teachers; "test scores do not necessarily show how a student can perform a related task in real life," and "parents and teachers frequently change their perception of a child because of a test score."

This group believes that school accomplishment is multifaceted, involving both processes and outcomes too complex to capture in a single score and must be interpreted in context to be meaningful. They plead for "authentic assessments"—those that are true to the local curriculum and pupil characteristics—and "real literacy" rather than accretion of separate skills that can be assessed in standard, multiple-choice format. Noting a connection between skills curriculum and the skills that tests such as the ITBS tap, they warn that schools impoverish education by adopting curricula that are consonant with mandated tests.

Although much of the political activity of CESE focuses on the ITBS, they also criticize mandated, centralized, and standardized criterion-referenced assessments.

FAIRTEST. As its newsletter states, FAIRTEST is a nonprofit "research and advocacy organization dedicated to ensuring that the 40 million standardized tests annually administered to America's students and job applicants are fair, open and educationally sound." Among its activities, FAIRTEST reprints and distributes articles in newspapers, professional and practitioner journals, laws and state rules and regulations, and decisions in court cases having to do with testing programs, results, reforms, civil challenges to test results, and interpretations of test results such as decline in Scholastic Aptitude results or differences between scores of males and females on items of the National Assessment of Educational Progress. In addition, its staff testifies before legislative hearings and acts as *amicus curia* in court cases having to do with tests.

Origins of FAIRTEST involved Ralph Nader and the public advocacy movement, and many of its officials are lawyers rather than educators. It receives funding from subscribers and various foundations, including the Ford Foundation. Although its interests are far-reaching, the issues that particularly energize FAIRTEST are the following: (a) ethnic and gender bias of tests such as the SAT, which result in inequitable allocations of educational and career opportunities; (b) effects of coaching programs on entrance tests that further detract from the already low validity of these tests; (c) misconduct and profit motives of large testing corporations, even those such as the Educational Testing Service that are officially nonprofit; (d) public disclosure of testing, score analysis, and score reporting practices; and (e) excessive and low-utility testing programs.

Based on content analysis of FAIRTEST materials gathered over three years, it is safe to assert that its beliefs encompass the discrepancy between indicators and traits of achievement. Specifically, test score differences between Anglos and minorities, males and females represent fallacious assumptions about mental processes and inadequate or racist and sexist methods for measuring them. Test scores are corruptible, by coaching or other means for preparing students to take tests. Coaching programs for the SAT, for example, reliably raise scores, but are not equitably available for all segments of society. Pressures to raise scores cause administrators and teachers to do what they have to do to raise scores, but corresponding effects on genuine achievement will not materialize. Tests harm children, teachers, and curriculum.

Cannell and associates. Cannell (1987), a physician in private practice, founded the Friends of Education after becoming aware of the contradiction between reported test results and alternative indicators of achievement. Why, Cannell asked, should all published reports of state-wide testing show most state averages higher than the 50th percentile on nationally normed standardized tests? He conducted his own study of state department reports and concluded that the public has been fleeced. Real achievement was substandard, while achievement test scores were high. Even southern states with abysmal records of SAT performance reported above average results. This he labeled the Lake Wobegon effect, wherein "all the children are above average."

Test publishers and others disputed his conclusions, claiming his results could be explained by the inferior methods he used, but a later replication with superior sampling and measurement supported his original conclusions (Linn, Graue & Sanders, 1989). Interpreting the effect he found, Cannell attributed it to two things: outdated norms on standardized achievement tests and outright cheating by administrators and teachers. In their desire to maximize profits, the test publishers go several years without gathering new normative data. Nor do the norms represent the nation as a whole. Furthermore, given a choice, school officials choose a test that makes them look the best[3]. Test publishers are "dumbing down" the tests; test items are easier now than they were 20 years ago. These things make it look like performance is improving, but it is not, says Cannell. In addition, tests are used over and over, so that teachers gain familiarity with the items and provide direct practice on those items, a practice that falsely elevates scores without changing the level of attainment. He believes that administrators, teachers, and test publishers collude to enhance the image of schools and defraud the public. They do this by cheating, pure and simple.

---

[3] In the reanalysis of the Cannell study (Linn, Graue, & Sanders, 1989) Arizona survived charges that all the states are above average. This may in part be due to the fact that unlike other states, districts in Arizona cannot choose the test that makes them look best (by Cannell's interpretation) or the one that best matches their curricula.

To Cannell, standardized achievement tests, at least before test publishers and educators started tampering with them, measured the trait of achievement. Negative publicity is the key to making schools work harder to increase real achievement.

Beliefs about Testing: Summarizing and Theorizing

Teachers define a substantial but not total discrepancy between the indicator and the trait of achievement. Their definitions of educational attainment are broader and at least in part inconsistent with models of teaching and learning embedded in achievement testing. They see up close what happens to test scores when pupils read poorly or lack facility with English, self-confidence, and middle class values of persevering in the face of frustration. They know from the evidence of their eyes and ears what happens when what they teach (whatever its merits might be) fails to conform to test content. Later, looking at the test scores, teachers can remember how hard the pupil tried or what else was happening to him or her. They can look at the score in relation to other indicators of achievement—daily performance, tests over material covered in textbooks, books read voluntarily, journal writing, conversations—and make a reasoned judgment about educational attainment, broadly defined. Thus, teachers have access to "interpretive context," that is, all of these other indicators, against which to judge the meaning of the score itself. Obviously, teachers are the only ones to have this interpretive context. Perhaps not all of them avail themselves of it.

When one compares teachers' beliefs with those of other groups, one finds the others more likely to assume a constant relationship between the trait and the indicator of achievement. The public views the relationship of the test score to educational attainment much as the relationship of yardstick to distance. Critics imagine a rubber ruler or ruler that works better for some groups than others in our society. Testing specialists work to preserve their status as the Bureau of Standards, more concerned with preserving the integrity of testing systems than with their effects on schools.

Although teachers value test scores for the information about achievement they convey, administrators seem to value test scores as organizational tools to reward, punish, cajole, and control, irrespective of the information they carry about real achievement.

Teachers have little use for the results of tests, although they believe, to justify its costs, that testing ought to be useful in advancing instruction or evaluating pupils. Nevertheless, they believe that someone else uses test scores, without benefit of interpretive context, against them: to shame them for putative laziness and ineffectiveness, to make them work harder, and to limit their autonomy to teach as they see fit. To defend themselves or in response to administrative directives, teachers strive to increase the numerical value of the indicator without regard to the effect on attainment. The means to increase scores are readily available: teaching test-taking skills and content they know is on the test. They disagree among themselves about what the effects of such training will be, but they concur that they have less time and energy left over to spend on education they value.

Chapter Three: The Natural History of the Testing Event

Introduction

To the casual observer, and perhaps to anyone outside the day-to-day life of elementary schools, the administration of the external test happens for a week in April and then is over. A few simple directions to pupils on how to use answer sheets, 90 minutes a day all week on the test itself, then back to the normal routine, the regular curriculum—this is how many outsiders imagine it. Those who work in classrooms, however, as well as closer observers of schools (as we may count ourselves by virtue of this study), understand that the testing event dramatically alters school routine before and during the test, and its effects reverberate afterwards. In this chapter, we wish to document the history of the testing event in two elementary schools and the roles testing plays in everyday life there. What we discovered in our analysis is that testing activities assume a kind of natural history, the stages of which are governed by the proximity of the external test. In each stage, patterns of teachers' actions and the meanings they hold change. Hence, the role of external testing changes across these stages, which are depicted in calendar form in Figure 1. Furthermore, the testing event is cyclical; the test results from one year are used to organize reactions to tests in the next year.

In Chapter Two, we offered our analysis of beliefs about testing. But beliefs are one thing, actions another. People do not always do what they say. Their meanings and intentions are sometimes more clearly understood by studying their actions firsthand and juxtaposing observations with their statements. Describing the role of testing in elementary schools requires a delineation of the actions of people within a social context. In this chapter, we take you inside two schools and describe everyday life there.

Following Erickson's (1986) recommendations about reporting qualitative research, we intersperse particular description, general description, and interpretive commentary. Our intent is to present descriptive data that are characteristic and typical of the cases we studied, as well as significant to the development of our assertions, so that the reader's thinking can follow the same paths ours did.

Testing Goes on Here

> Forget your nostalgic memories of grammar school. This is Hamilton Elementary School. Testing goes on here, but testing is not the first thing you think of. When you first see the building, you feel that it is of a piece with the surrounding neighborhood. Not too clean, the paint peeling in places, its drabness owing to its place in the district refurbishing cycle. The "finger plan" that makes up its campus reminds the visitor of old army barracks. Two rows of separate classroom buildings are linked by a covered breezeway. Several permanent looking "temporary" buildings house special programs like Head Start and instrumental music.
>
> A convenience store across the street serves a dangerous-looking clientele: heavily tattooed bikers and drivers of old pickups with gun racks or chopped and channeled Chevies. Teachers encourage us to avoid the store,

Figure 1

CALENDAR OF ACTIVITIES IN THE NATURAL HISTORY OF THE TESTING EVENT

PRIMARY ACTIVITY	SECONDARY ACTIVITY	TYPICAL DATE
REACTING	FORESHADOWING	August
ORGANIZING SCHOOL		August, September
PUTTING TESTS IN BACKGROUND, ORDINARY INSTRUCTION IN FOREGROUND		September
PLANNING FOR TEST	CONTINUING ORDINARY INSTRUCTION	January
PUTTING ORDINARY INSTRUCTION IN BACKGROUND, TESTS IN FOREGROUND		February
TESTING	PREPARING FOR NEXT TEST	April
RESTING	REORGANIZING SCHOOL	May
REACTING TO SCORES		June
ALIGNING INSTRUCTION	FORESHADOWING	August

and suggest we eat our lunch at school and park only behind the school in the fenced parking lot. But this is no ghetto or barrio school, and other things about Hamilton compensate: the classrooms are cheery and welcoming, the people are friendly, and there is a huge vegetable garden that the children delight in working. A large sign with movable letters stands in front of the building, adjacent to the office. "Free education," it says today, "Bring your own container."

Looking at the houses and apartments near by, you see windows boarded or with bars, derelict cars lying on their axles in the dirt. Familiar urban characters with bags and shopping carts can be seen picking through the trash lying about, the abandoned upholstered furniture out on the street, the weeds growing in vacant lots contrasting with the blooming bougainvillea next door, the dust of Phoenix drifting in near-constant sunshine. Fierce-looking dogs in the backyards near the playground; Dobermans and rottweilers are the breeds of choice. You see some graffiti on concrete block walls, but gangs and crack houses concentrate in other parts of the city. Houses and yards vary in how well maintained they look, but keeping up with the yuppy Jones' is not what drives the inhabitants of this neighborhood.

Some of the teachers say they refuse to hold conferences with parents in their classrooms at night. One female teacher brings her husband along. We never went to school after dark, but perhaps others would not have been so squeamish.

Like most urban cities, Phoenix is full of contrasts. Here the rich and the poor neighborhoods are often separated by only a high stucco fence, a canal, a hedgerow of oleander, or what passes in this desert for mountains. The patchwork of elementary, secondary, and unified school districts in the metropolitan area, and the schools within them, are gerrymandered in ways incomprehensible to the outsider. Some are small and uniformly minority and poor. Others, like Cactus, are large, relatively well-off, and diverse. Although the trend is there, wealth has not completely escaped to the Phoenician suburbs, but concentrates in neighborhoods and school catchment areas. Hamilton is not unusual in the metropolitan area in being a economically poor neighborhood school in a middle class district with traditional American values, ambitions, and images of itself. Hamilton stands out from the the district norm, though two other schools, including Jackson, come close. Hamilton has no middle class group to elevate its average achievement or social tone. When the kids graduate to junior high, they will come face-to-face for the first time with polo players or alligators on the clothes of their classmates. Now the most trendy items of clothing you can look forward to seeing are t-shirts emblazoned with graphics of Spuds MacKenzie or the heavy metal rock group, Metallica. The kids mainly wear blue jeans or athletic shorts or sweat suits, the girls the same as the boys. Some will be meticulously groomed, others will have greasy hair cut in amateurish ways.

But now, on August 20, it is still too early in the year for the students to appear. Only the staff comes in early to get organized, set up their classrooms and attend meetings. When the students do arrive, nearly 800 are expected, and they will also reflect the community: 75 percent will be Anglo, 12 percent Asian, 10 percent Hispanic, the rest Black or Native American. Asian immigrants, as well as immigrants from Central America and Africa, stop over in this neighborhood, at least until members of the family find jobs. Then they move on to better neighborhoods and other schools. The less successful ones, so the teachers say, stay here.

Ask the teachers what their biggest problem here is and almost all of them will tell you about the transience, not just of the immigrant children but all of them. It seems to the teachers, though, that the children who come in are not a random group, but perform near the bottom of the class. The kids who move around are those whose parents can't make the rent so they have to move, or their families break up and regroup, when movement is viewed as a solution. The kids move from school to school, even within the district, with few—and delayed—paper trails. For some of them, they will enroll in five or six different schools in a single year. Hamilton's official rate of outward mobility for last year was 32 percent, with 15 percent coming in to replace them. It is commonplace for Hamilton teachers to finish up the school year with only half of the children with whom they began. This is a fact of school life that worries most of the staff. Teachers believe themselves effectual—if only they could have some continuity with their students.

The most over-subscribed program in the school will be the English as a Second Language (ESL) program. Students whose native language is something other than English and who pass certain other criteria, will be pulled out of their regular classrooms for one, two, or three hours each day to learn English. Most of them are successful eventually, but it takes some of them more than a year. The ESL teachers guard their load carefully. If they were less vigilant, they would be given twice as many students as they could handle. The regular teachers believe these immigrant children need more than a year in the program even when the official rules declare they are no longer eligible. The non-English speakers often mark time in their regular classrooms, where instruction goes on without their participation while they draw pictures or simply sit and wait for their assigned time in ESL or for the end of the day. Conversations among pupils heard after school is as much a polyglot as you are likely to encounter anywhere outside the Heathrow customs area.

Sixty percent of Hamilton's pupils are eligible for free lunch, a mark of poverty, the highest rate for any school in Cactus District. Many come for free breakfast as well. Parents are recruited to help serve these meals, but parent participation in other activities, like the Parent Teacher Organization, is almost nonexistent. When parents come, the event that is likely to draw them is a cookout or some entertainment where students perform. To call Hamilton's population working class would be too optimistic. Many fathers don't work at all, or work periodically, get laid off, accept unemployment or welfare, or move in with family members who can support them for a time. Some fathers no longer provide for the children. Mothers work and leave their children to their latch keys after school. With economic instability often comes family instability, according to the teachers. Parents desert or divorce, or never were married to begin with. Single mothers take up with boyfriends, who, like stepparents, are sometimes less committed to the children. Some have more commitment, and become the source of strong emotional attachment without the long-term security that makes such attachments trustworthy. High on the list of concerns of teachers are the abuse and neglect they believe are commonplace in the children's home life, the lack of concern for education that they read in the parents' reactions, the many times when children report they went to bed late or were not given breakfast, the poor habits of nutrition, cleanliness, and health care, the hostility that occasionally erupts among students, the head lice, the access to drugs.

It takes a special person to work here. Despite the many problems, teachers say they wouldn't want to work in a middle class school. Here, they say, there is a chance to make a difference in the lives of children. They have started many special programs to meet the special needs apparent here. To the staff, the school is the best thing in the lives of these children, an island of stability and order in the midst of chaos.

But, as obvious as the teachers' caring for individual pupils, the relationship between staff and community is at arm's length. Phrases such as, "The kind of kids we have here..." or "Pupils like these..." often preface justifications for programs at Hamilton and determine a kind of image the teachers have of the children and their families.

Jackson Elementary School is only a few miles from Hamilton and resembles it in many respects, starting with its size and the layout of its campus. It has been renovated more recently, however, and looks brighter and cleaner than Hamilton. Its population is also primarily lower income, lower middle class, but has within its boundaries a few middle-class town house developments. The surroundings of Jackson are notably better kept up and safer (though many of the teachers would not agree). Its rate of transience is 28 percent, compared with Hamilton's 32 percent. It has fewer ethnic minorities than Hamilton (82 percent Anglo, 12 percent Hispanic, 3 percent Black, 2 percent Native American and 2 percent Asian) and few ESL pupils. It's rate of free lunch recipients is 33 percent (compared to Hamilton's 60 percent). Unlike Hamilton, Jackson has an active group of about a hundred supportive parents. Though Jackson is not as poor as Hamilton, Jackson's teachers express similar sentiments about the differences between themselves and their community, and similar beliefs about "pupils like these."

In these two settings, then, we confront the external testing programs in the real world. These are, perhaps, the schools of the future, the near future—almost close enough to touch.

Stage One: Reacting/Foreshadowing

Even now, on August 20, 1987, the prospect of external testing rumbles in the background of teachers' thoughts, as they otherwise devote their attention to the nuts and bolts of school: setting up their classrooms, collecting texts and materials, checking and juggling their rosters, and attending district, school, and grade level faculty meetings. Among other things, principals distribute all the test booklets and answer sheets for the formative and summative tests that the district mandates for the upcoming year. The new fourth-grade teacher, for example, would be receiving three sets of Continuous-Uniform Evaluation System (CUES) in reading, language arts, and math, Basic Skills Tests (BST) on the same subjects, plus science, social studies, and study skills, and an elaborate district Scope and Sequence. In Cactus District, teachers know that pupils' test performance is important, and they know as soon as they see these tests what material they should cover, and in what format to cover it. Targeted instruction—that is how the District perceives proper teaching. Not every teacher is on the same page—this District is not so uniform—but every teacher is pursuing common goals, and is accountable for the same standards and evaluation instruments.

Although the target is far off, principals bring testing into the teachers' thoughts during this first week's meetings by reviewing last year's school performance on tests, setting goals for the coming year (some of which refer to the school's performance on external tests), and translating the messages principals have heard from their own superiors. Later in the week at a meeting for all the teachers in the district, teachers will listen to inspirational messages from superintendents and other district administrators about their vision for the district and their expectations for the teachers. The history of the testing event has begun.

Take the experience of teachers at Hamilton School. By the time of the first staff meeting, teachers will have attended five or six meetings at the district office or in the school. They will have received special training in, for example, the district's pilot program in kindergarten math or the district's new writing program, which bears the title of *Writing and Thinking*. Some will have gone to meetings of special education case coordinators or the district reading textbook adoption committee or science building coordinators. Besides seeing to these activities, common across the district, Hamilton teachers attend in-service training in various Direct Instruction packages: *Reading Mastery* (the "heart of our school," says the principal), *Spelling Mastery*, and *Expressive Writing*, all of which require teachers to follow exactly the scripts and uniform procedures in the manuals. In addition, teachers must learn the procedures of the discipline program that is common throughout the school, known as ATF, or Attitudinal-Transitional Format. This is a program based on behavioristic psychology that the district has used for a number of years in its self-contained program for emotionally disturbed children. The principals adopted this program at Hamilton because of its large proportion of disadvantaged children and because of its earlier reputation as a school out of control, a "real zoo." To implement this program successfully requires that every teacher follow the specifications exactly the same as every one else, so in-service training seems necessary.

Besides the in-service training, the principals will make sure teachers understand and follow the common programs by observing and evaluating each teacher about every two months throughout the year. The principal or the district's ATF coordinator will come into the class, take detailed notes on classroom transactions and teacher's behavior, and then conduct a conference with the teacher about his or her strengths and weaknesses in following the programs. The principals credit this method of supervision to the Elements of Effective Instruction (EEI) or Madeline Hunter program, which many Arizona administrators endorse and use. Anyone new or in need of a refresher course will have to attend an EEI in-service before the start of school.

There are obvious themes to Hamilton's opening of school. Dr. Thorne, the principal, Dr. Michael, the assistant principal and in-service trainer, and most of the teachers have strong intellectual commitments to Direct Instruction models of teaching. They believe that when teachers implement these models correctly, pupil achievement (broadly defined and defined by the external tests) will prosper. This view of testing and the function external tests serve during this stage of the history at Hamilton play out at the opening staff meeting.

Opening Staff Meeting at Hamilton

> Dr. Thorne assumes his place in front of the long shelves in the school library where he can see the faces of most of the staff. There are 47 people present, about one-quarter of them male. People are dressed casually to ward off the August heat of Phoenix, some sitting at tables and others lounging on overstuffed furniture. The mood is light, friendly, informal. Dr. Thorne greets different ones, sharing jokes and stories. Although you could pick him out of the crowd as the principal from his dress and appearance, you couldn't distinguish him by his manner of treating people. He is professional and personable, with no boss-worker tone, and teachers seem to respond to him. There is no sign of the familiar scene in school faculty meetings where a cabal in the back row smirks or sleeps.
>
> He opens the meeting by noting the change in superintendency in the district, and the effect of this change on Hamilton. "The new superintendent's philosophy is that individual schools will have the license to operate to meet the needs of the kids at the local level." New-found autonomy is a recurrent theme, already sounded in Dr. Thorne's welcoming letter to the faculty: "This will be an advantage to [Hamilton] staff and students since we are well on our way to doing 'our thing.'" He announces other changes new this year: teacher evaluation will be different, and computer literacy is a new district goal they will pursue.
>
> "What will not change from last year is the expectation for good teaching." The major programs from last year, Reading Mastery and the ATF model, will continue. They will add Distar language, spelling, and writing. Research bears out the effectiveness of the Direct Instruction program, "particularly with the kind of kid we are working with. We all need to pay more attention to educational research." He distributes reprints of "What Works," the U.S. Department of Education document that extols the merits of phonics-based reading programs, and copies of research studies that show the success of Direct Instruction programs.
>
> Proceeding around the room to introduce the teachers, Dr. Thorne incorporates personal comments about each one, where they previously taught or went to school, recent marriages and children, and humorous incidents that occurred when he took a group of teachers to California to attend a conference on Direct Instruction. At that conference they talked to teachers at other sites carrying out this program, but unlike them, Hamilton has "complete administrative support and the best possible training, conducted by Dr. Michael," who has authored several Direct Instruction programs. Again, Dr. Thorne draws teachers' attention to

evidence from achievement test scores that proves the success of Direct Instruction. In this way he moves the meeting to a new topic, his presentation of last year's results on the Iowa Tests of Basic Skills.

"Looking at the rank order of schools on reading comprehension," he says, "there was only one grade level where we had less than one year's growth. We exceeded the average or norm of reading comprehension of all the other schools. This signifies real good teaching and learning. This is big stuff. This is significant. On language, we were outperformed by only three schools and those kids [from more privileged neighborhoods] are in a different world! In math this school did exceedingly well, above the mean; on study skills, we did a good job in that area too. The report shows that we focused on instruction and making kids successful. And the district sat up and took notice. We got their attention." The teachers listen politely, but offer no comment or question. They have no written report to follow, but the message seems palatable: that the test scores are valid indicators of quality schooling and that desirable scores are attainable at Hamilton with the current programs.

Shifting topics again, Dr. Thorne outlines the goals for the year: continue the reading and language programs and the ATF discipline model, initiate computer education, continue the extended day program through homework club and Saturday School, increase oral and written communication and problem-solving skills, increase parent involvement in the schools (there are presently only two or three parents involved in PTO), conduct education about substance abuse and increase attendance. "The big push" will be Making Your Day, part of the ATF program that rewards good behaviors.

He continues through the meeting's agenda, spending time on the details of the ATF program, book orders, appointment of instructional leaders and grade level chairmen, the garden program, and the administration, scoring, and reporting of CUES testing. When someone asks how CUES will be handled this year, a *sotto voce* comment, "Very poorly," draws chuckles. One gets the idea even now that teachers have little respect for CUES, even though they will have to administer them three times during the year.

Dr. Thorne turns the meeting over to Dr. Michael and the teachers break into applause when he introduces him as the new assistant principal. Dr. Michael turns the topic of the meeting again to testing and the results from last year. Using the library shelves to illustrate achievement levels, he explains, "Our kids started down here and made progress up to here. Other schools may have started higher than we did, but they didn't make as much gain in a year as we did. In absolute terms, we started lower, we made better gains, but the others are still ahead of us. What is our goal? To get all the kids up to the national mean, median, or average. Fifty is the magic number. But you see that a lot of kids didn't make it. You know who they are, and I know who they are. We have a large group of high ability kids, but the performance of the low ability kids brings down their scores into this average. Research shows that the low income kids are around the 16th or 17th percentile."

> Turning to the results of the Metropolitan Achievement Test, an internal test that the principals have elected to administer in April, he reports, "On the Metropolitan, the overall average of the first grade [total reading score] is at the 47th percentile. But if you take those kids who were here 120 days or more, the average is at the 51st percentile." He shows that if you exclude from the average those children who are in the transition class, the average of the group in attendance 120 days or more is at the 57th percentile. He notes that the averages are higher in the upper grades; the average on the Total Reading scores for sixth graders who had been in attendance 120 days or more and were not in the transition class was at the 65th percentile. Reporting this difference between the performance of primary and intermediate grades, he says, is not meant to be a criticism of the teachers in the lower grades: "It's praise. It shows that what we are doing is working. We are off to an excellent start." The primary grade teachers are sowing the seeds and the intermediate teachers are reaping the rewards."
>
> Dr. Michael ends his presentation of last year's results with information about upcoming tests. He encourages the teachers to do as well as or better than last year. This is all quite straightforward and businesslike, and the teachers seem to take it in the same spirit. There is no open questioning of the interpretation of the results or the wisdom of testing. The only hints of disagreement come outside the meeting. During a break, a primary teacher calls the ITBS and the Metropolitan "cruel and unusual punishment" for pupils in primary grades. This was the opening note in a recurrent theme about the perceived deleterious consequences of testing on pupils.
>
> Ending the meeting is Dr. Thorne's message of encouragement and competition. He speaks of the many awards and recognition that some other schools and districts garner, attributing their success to their "teams of writers" and public relations specialists rather than their superiority in programs. "There is nothing they do that is any better than what we do. We're going to be a leader in this district. This is not just some little local thing that we're doing here." Speaking of the "A+ Schools" or "Top Ten Schools" awards, he says, "Even though our kids function lower, we're going to go for it. We're an effective school. What the nation is talking about regarding what's effective is what we're doing. This will have national significance."

When the principals at Hamilton review previous ITBS performance, they typically draw the attention of teachers and others away from the familiar percentile ranks and grade equivalent scores and toward what they define as the "gain score." This is not the definition of gain score that psychometricians would recognize, because it is not the difference between scores obtained by particular pupils attained in second and third grade, averaged across the number of pupils. Instead, the group-gain calculated by Cactus District is the difference between the average grade equivalent score of the third grade in one year and the average grade equivalent of the school's second grade the previous year. Because of the high rate of pupil turnover in this school, approximating 50 percent, the grade equivalent score for the third graders is made up of no more than half the individual pupils from the previous year. When he states that Hamilton exceeded the average of all schools on reading comprehension, Dr. Thorne refers to the group growth score and not the average grade equivalent score. Hamilton's ITBS reading scores in 1987 were lower than grade placement in three out of six grades, but group growth exceeded the district standard of one year in grade equivalents in all but one grade. This year, at least, Cactus District is stressing the importance of the group growth score as evidence of school and principal accountability. To the principals and teachers of low-scoring schools like Hamilton, use of the group growth offers a ray of hope and raises the possibility of fairness in accountability. Although they often despair over their chance of bringing their pupils up to the national average, they feel their efforts might be more fairly judged by looking at the progress their pupils made. Most feel that using the group growth takes into account where the pupils started and how much progress they made in a year. Although they express some doubts about the effect of transience on the meaning of the group gain, they do not seem to be aware of the plausible alternative interpretations one can draw from such a difference.[4]

Group growth is not the invention of the district. The Arizona Department of Education uses a similar statistic as an indicator of progress in tested achievement, reporting what they define as "growth" as the difference in percentile ranks between, for example, a school's third graders in one year and its second graders from the year before (Bishop, 1988). In the state's analysis, however, they make some attempt to match up the populations tested in the adjacent grades.

Besides the group growth on ITBS, principals at Hamilton have another way of showing the success of their program. On their own, not by state or district mandate, they administer the Metropolitan Achievement Test. Besides being more in tune with the Reading Mastery program, the staff can administer the Metropolitan reading test within functional reading levels rather than within grade levels. That is, a sixth grader who is reading at the fifth grade level in Reading Mastery would take the fifth grade form of the Metropolitan. The state requires, however, that the same student take the sixth grade form of the ITBS. As Dr. Thorne says, "These kids are savvy. They know what the game is. We have them [sixth graders] working successfully on Level 4; then they have to take the sixth-grade test [ITBS], and they fail. It reinforces the notion that they can't do sixth-grade work." According to his view, the Metropolitan is more appropriate. Dr. Michael is conducting his own research on the success of Reading Mastery using Metropolitan test scores.

Because the Hamilton principals score and report the Metropolitan tests themselves, they analyze the data in ways meaningful to them. They present scores separately, for example, for groups of children who had attended 120 days or more and those who had been in school fewer days. For them, the 120 day cutoff was an arbitrary figure, approximating two-thirds of a school year. "We feel that if we can teach them that much, that is the kind of effect we can have." Like the group gain statistic, the breakdown of scores by attendance is consistent with the teachers' notion that pupils' attendance is somehow out of their control. They feel that they should be held accountable only for those who attend or whose parents demand their attendance.

---

[4] See Cook and Campbell (1979) and Cronbach and Furby (1970) on the unreliability of measures of change and alternative interpretations of changes. The district's interpretations of growth scores as indicative of adequate teaching (and alternatively, their interpretations of less than a year's gain as indicative of inadequate teaching) overlooks problems of the poor reliability of gain score generally. The difference may be caused by maturation differences in the group, nonequivalence of the two samples, practice effects of testing, statistical regression, and nonequivalence of samples. Even more obvious threats to the district's interpretations are sample attrition (different individuals being tested at the two points) and instrumentation (different test samples and forms administered at the two points and different raw score distributions test publishers use to compute grade equivalent scores at different grade levels).

Dr. Michael also reported Metropolitan scores separately for transition and regular classes. The transition classes are a major structural element in Hamilton. Instead of having special education classes and teachers, the staff groups "mildly handicapped" children with children who are well below their grade level to form a transition class. These classes are small, and students progress through the normal Reading Mastery curriculum at a slower pace than regular classes. Most are taught by teachers certified in special education. There is a transition class for each grade, first through sixth. Excluding scores of children in transition classes from the average Metropolitan test score results is congruent with the teachers' ideas that test scores reflect the raw intellectual abilities that are outside the teachers' sphere of influence.

A short distance away, at Jackson School, the nuts and bolts of getting ready for school run much to this same form. Yet the rhetoric about last year's test performance and this year's testing schedule differs. By his own admission, Dr. Thorne plays "the game" of test scores and works to enhance his school's chances of winning it. Mrs. Mitchell, principal of Jackson School, openly rejects tests and the use of test scores as antithetical to the school's philosophy. Mrs. Mitchell calls Jackson a "Whole Language School." Teachers transfer to Jackson because of Mrs. Mitchell's reputation as a Whole Language Specialist. Not every teacher follows her philosophy, but those who do practice some other instructional method also participate less in the school's rhetoric. The dominant view is that standardized tests are contrary to the Whole Language philosophy of education. What those tests cover is not pupil understanding but meaningless bits of knowledge and skill. Tests overemphasize comprehension of sentences and paragraphs, isolated from meaningful context in real texts and authentic communication. In a series of meetings that first week, the staff addresses the discrepancies between tests and learning and attacks head-on the district's use of test results to judge schools.

Opening Staff Meetings at Jackson

> Introducing, inspiring, reviewing, organizing details, and projecting into the future, Jackson's initial staff meeting is not unlike Hamilton's. But when Mrs. Mitchell comes to the agenda item on the ITBS, the coverage is no more than cursory. They all know that the outside world will judge them on their test scores, and they expect those scores to be low. In this vein, Mrs. Mitchell reports on a conversation she has had with the assistant superintendent. In that conversation, he had expressed his support of the Whole Language Program and downplayed the tests as valid indicators of their program. He hinted at lenient treatment of Jackson's test scores. When she had asked him about Jackson's possible use of *Scoring High on the ITBS*, a test-preparation program, he discouraged it. He suggested that less friendly audiences would attribute any good scores that Jackson might attain to their use of *Scoring High* rather than to the success of the school's program.

Some of the teachers are skeptical about this secondhand report. What are his real motives, they wonder? Does he really expect and even want them to fail? Will low test scores result in the withdrawal of Jackson's license to practice Whole Language? Could they even lose Mrs. Mitchell as principal? Mrs. Mitchell tells them that she will invite the assistant superintendent to come and address them on these issues.

Like Reading Mastery at Hamilton, Jackson's use of Whole Language amounts to a kind of variance from the district's prescribed curriculum, scope and sequence, and basal series (Ginn) used throughout the rest of the district. Both schools operate as pilot schools for their various innovative programs. Many of the teachers perceive that low ITBS and district test scores will result in the district withdrawing of these variances.

In the initial meeting, Mrs. Mitchell makes only sketchy and superficial use of last year's test scores. She asks that during the grade level meetings the teachers "take a look at" the scores from last year, particularly the group gain scores, and "see what you can make out of these wonderful things." She tells them that she analyzed the scores from only those pupils who had been at Jackson for all six years, "because with tests there are so many variables." She thought that these scores might show that the Whole Language program had made a difference, but they did not. "I don't know what that means." One teacher says that they need at least three years to demonstrate a difference, that the tests scores are simply not sensitive to a single year's effort. Mrs. Mitchell also provides copies of CUES and BST tests they will have to administer during the year. She asks them to look at the material covered on the tests and "think about units," or integrated lessons and projects that will accomplish the goals the district has set. They oppose the practice of using the standard texts and formative district tests to determine what is taught. Instead, they study the texts, Scope and Sequence, and the contents of tests for the goals and objectives the district values, then incorporate these into units such as mythology or ecology or urban geography around which they organize the teaching of literature, writing, and content knowledge.

Mrs. Mitchell announces that Jackson will be a pilot school for a new, experimental form of CUES testing in language arts. Instead of a paper-and-pencil, multiple-choice format characteristic of most CUES, the pilot form involves teachers' ratings of accomplishment of the desired competencies; the teachers respond enthusiastically to this.

During the grade level meetings, only the first- and second-grade teachers address the topic of external testing. Specifically, they discuss parents' lack of understanding of test results, the discrepancy on the part of some pupils between their low ITBS score and their real ability to read, the low probability of Jackson's ever attaining scores as high as some of the middle class schools in the district, and the harmful effects on young pupils of taking the test. Teachers express considerably more interest and concern for grading than for testing. Grading and testing are completely different activities, and external tests and even district CUES and BST have no utility for them when they must assess pupil accomplishment and assign a grade. Even for first graders, the district requires teachers to assign grades, and the process of grading is the province of the individual teacher, although teachers within a grade level share their own procedures and attempt to agree on a common plan. Thus, in the grade level meetings and the faculty meeting the teachers grapple with different grading schemes and the controversies over each one. These discussions fade without reaching a decision. No one seems concerned. It is clear, however, that the ITBS and BST results are meaningless for this task. The CUES merely take time and paper, without returning much utility for grading.

One week later, the faculty assembles for a second meeting. There are about 40 teachers present, and everyone is dressed comfortably and seated at round tables. Coffee and donuts are on an adjoining table. Mrs. Mitchell presides, proceeding not quite directly through the agenda covering rules, procedures, and budget for materials, sprinkling the dry stuff with supportive and encouraging comments. She announces that the assistant superintendent, as she promised, will come to talk to them sometime during the meeting. One teacher asks, "Is it okay to ask him about tests?"

Mrs. Mitchell announces that the three district priority goals are drug abuse, problem-solving, and absenteeism. "But these things have to do with what is happening in the classroom. In our program and the self-contained classrooms, we have the time to air feelings, we work on self-esteem and decision-making. Doesn't this relate to drug abuse...?" She is saying that Whole Language gives them a reason to be in school, and direct dealing with choices and feelings may give them reason to avoid drugs. She also says that these are important goals, but "we have a broader purpose."

As she covers what is in the teachers handbook, she tells them that everyone must conduct the student survey on reading. They are doing the survey in the fall "as a pre-post thing, because we're going to need some kind of data to show what we're doing is working." The survey is about reading habits and attitude toward reading. A parent survey is also part of this "thing," although, instead of pre-post, it is to be Jackson compared with some other school in the district. By doing the survey, Mrs. Mitchell is seeking an alternative to standardized test scores as sole indicator of school effectiveness. Consistent with the school's philosophy, the survey will measure whether the child "is a reader," whereas tests like the ITBS measure fragmented reading subskills out of context with instruction, according to Mrs. Mitchell.

At this point the assistant superintendent comes in. He seems warm and sincere, and the teachers are polite and attentive. He speaks ritualistically about a great new year and adds, "We need a great year. Due to the new superintendent, there is a real feeling of growth, of wanting to see how we can do business better. We want to empower people who are close to the students, toward greater decentralization....It's important to process a longer term view than just one year. Here we have a whole fresh group of kindergartners, full of hope at the beginning of the 13 years they will spend in the district. They will graduate in the year 2000. Where will you, where will all of us be in our careers in that year? And, how do we define what ought to be the criteria for success in these 13 years? How would we like to get them ready? I'd like to see them have as many options about what to do with their life as possible. I'd like to see how we can engage in planning to make that happen....We need to think ahead about the decisions we make. For example, what does the decision mean when we decide a kid is LD? Does it mean that he won't experience a curriculum that eventually could get him into college?"

"I believe that success builds success; that all kids can learn, the SAME curriculum as anyone else; that good schools accept the responsibility for control over conditions that let our kids be successful. Sometimes this is a differentiated curriculum, sometimes different methods, sometimes MORE TIME. We need to learn how to manage time."

"What's exciting about Jackson, under Mrs. Mitchell's leadership, is that you've asked for flexibility, for empowerment so that you can create the conditions for learning. You may make mistakes, but you will have the responsibility and the trust."

A teacher says, "A lot of us have the philosophy that tests are not conducive to learning. We think we should have a three-year trial period when we didn't have to worry about keeping scores up. We feel pressure, not necessarily from administrators, but from other teachers to do well on the tests the legislature requires."

He replies by asking how the rest of the teachers feel and gets consensus. "I'm cautious about tests. We need to realize that not every test is a standardized test. Evaluations are something else. I believe we have to do it. I believe we have too much standardized testing for the good of students. But I haven't been asked to do anything about this [laughter]. Many people like Mrs. Mitchell go down to the legislature every year to testify against more standardized tests, but the legislators aren't convinced. But because of your program, we'd like to give you some leeway. I don't think you should have to worry so much about your test scores."

Another teacher comments on the absurdity of giving grades in primary science. He says that because of the nature of the program you can develop other methods of grading. Also, if there is no basis of grading in a subject, that it is not necessary to give a grade.

The teachers express concern that their program is unique and the district should not use test results as the sole basis of judging its merits. The Assistant Superintendent responds: "The key word in the administrative staff goals is responsiveness. The best protection for your program is your community's support. Parents won't allow us to take away a program from their kids that they value. The key is parent involvement and responsiveness to it." He closes with a story and a parable about trust, responsibility, mistakes, and more trust. "I want you to have the same autonomy."

He leaves, and the teachers break into applause that seems quite sincere. Yet the sense remains throughout the year that, in spite of his good intentions, the district will use their test scores and jeopardize their program. This sense grows as teachers get the word from others throughout the district that central administration has set up an incentive system for principals: dinners for two for principals whose school show more than one year in group growth scores on the ITBS, reduce absenteeism by a set amount, produce the best writing assessment plan, enlist the most parent volunteers, or win the "A+" award. Later in the year, when it becomes known that ITBS and BST gains will determine part of principals' evaluations and merit raises, the defensive stance taken by Jackson teachers solidifies.

Testing at Stage One. At these schools and perhaps others as well, teachers and principals start out the year with a set of philosophical commitments to special programs, even those as disparate as Reading Mastery and Whole Language. They may have taken an interesting class in summer school or heard of a new teaching approach; the new year brings new hope and a sense of opportunity. Perhaps they want to try out something new: a new way of teaching math, new stories they have found; maybe they are ready to try a problem-solving model of instruction. But aside from these personal goals, facing them in the year ahead is the prospect of carrying out an array of activities and special programs from drug abuse prevention to computer literacy. The stacks of district goals and texts—health and spelling and writing and language and reading and Arizona history and new math and old math—remind teachers that the district expects them to cover a daunting body of material. How can it all be done? Now, even in August, teachers wonder if they will make it. Later, when the demands for work exceed the time available, they will make choices. Some things will have to slide. The role of testing at this stage is to suggest a priority to teachers about what can safely be omitted or neglected in favor of covering content they already know the tests will cover.

Recitations of last year's test scores and reminders of what happened as a result of the scores set in motion a series of actions by staff to avoid those consequences and public failure the next time.

Stage Two: Opening and Organizing School

At each school, tests play a role in how pupils and teachers are assigned and organized into groups. Because we have been raised in the same system, most of us are apt to take for granted that pupils are put together somehow into collections, and this process of grouping is never random. Even the notion that pupils are organized by grades that roughly coincide with chronological age is a matter of school structure. The most extreme use of tests to determine school structure would be administering tests such as achievement, readiness, or IQ tests, establishing cutoffs, and assigning pupils to groups based on their scores. Tracking by measured ability, placement of pupils in between-grade transition classes, identification of pupils as eligible for special education or programs for the gifted are all mechanisms by which educators use tests to structure schools.

After several years' experience grouping pupils by ability and teachers by specialty (i.e., science teachers), Mrs. Mitchell and her staff at Jackson are trying out a self-contained organization structure: the principal assigns a set of pupils to one teacher who teaches all subjects, except for "specials" (music, art, and physical education), which the district requires that specialists teach. There is no switching between teachers (e.g., for science). Jackson takes this pattern further by reducing pull-out programs. In the typical pull-out program, handicapped or ESL pupils leave their regular classroom for an hour or two during the day or week and go to a specialist's classroom for instruction. For the rest of the day, these children are "mainstreamed," in regular classes. At Jackson, in contrast, learning disabled and other mildly handicapped pupils, those with limited English, speech defects, or poor reading skills, stay in their home classes. The specialists, including LD, Chapter One, ESL, speech, and "Skills" teachers, come to them, often sitting with the children in a corner of the classroom and working on the ordinary academic work that their homeroom teacher has assigned. Thus, there may be more than one professional teacher in a classroom at one time. The rationale for this structural arrangement, according to Mrs. Mitchell (this was her idea), is to enhance the sense of community in the classroom and preserve the teachers' sense of responsibility for all the pupils without exception. Another effect of this arrangement is to decrease the ratio of pupils to teachers when the specialist is in the classroom. For Jackson's teachers, this is a novel idea. Some doubt the wisdom of so much coming and going by specialists and worry about having another professional in the room. Classroom teachers and specialists struggle with this uncommon arrangement early in the year; later most will come to terms with it.

One pull-out program that remains part of Jackson's structure is the program for the gifted and talented. At different times during the week, the "Project Potential" teacher will appear at the classroom door and the "gifted" pupil will pick up his or her materials and leave for an hour. Like the music, art, and physical education programs, Project Potential seems immune to the principal's tinkering. Teachers in these categories appear to serve interest groups outside the school itself and proved the least compatible with the character of the school. This is true of both schools.

At Jackson, Project Potential is the only part of school structure where test results determine group membership. Acting according to state guidelines, the Project Potential teacher combs the ITBS printouts for pupils with scores above a certain cutoff. When she identifies these children, she requests that their teachers excuse them from class to take the Cognitive Abilities Test, which has verbal, quantitative, and analytical components. She assigns to the program those pupils who attain a score above the cutoff. These children will then pursue an individual curriculum that the teacher develops and manages. In some cases, children she identified in one year are simply carried over into the next. There is allowance, in state guidelines, for teachers to recommend pupils they deem gifted, thus overriding the testing procedures. According to some teachers, however, the specialists actively resist including those children who score lower than the cutoff but seem gifted by the definitions of the teachers.

In other categorical programs, such as learning disabilities (LD), disadvantaged (Chapter One), and English as a Second Language (ESL), tests play a major role in determining which children receive special service. These tests include the Weschler Intelligence Scale for Children and various perceptual-motor assessments for children whose teachers suspect are learning disabled; ITBS, informal reading assessments, and Reading Miscue Analysis for Chapter One programs; and the language tests for children with limited English. State and district guidelines primarily determine the identification procedures for these categorical programs, but often the staff at Jackson modifies the decisions that test results would ordinarily trigger. For example, Mrs. Mitchell announces at the initial faculty meeting that she wants no kindergarten, first, or second-grade teachers to refer their pupils for LD testing. Instead, she feels that teachers need to adapt their own programs to deal with children of many abilities, including those whom some other schools might label LD. Jackson uses the TAP procedure (Teacher-Assisted Planning) in which the principal, psychologist, social worker, LD teacher, nurse, and teachers convene to discuss particular pupils with whom the teachers have difficulty. Although most schools treat referral to TAP as preliminary to a special education staffing and perhaps as a means of removing a pupil from the teacher's responsibility, Jackson treats TAP as a way of providing support for teachers to develop alternatives for children in the classroom.

Children with serious handicaps receive their instructional programs in self-contained programs in special schools. Therefore, one is not likely to find blind or mentally retarded children at either Jackson or Hamilton. According to district guidelines, specialists administer tests to children such as these and place them elsewhere in the district.

Tests determine the internal structure of Jackson in another respect. Kindergarten and first-grade teachers decided to provide a transition first grade for last year's kindergarten children who fell below a certain level of readiness on the Gesell readiness scale. This year's kindergartners are given placement tests, the district prescribed Learning Accomplishment Profile, Diagnostic Kindergarten Profile (LAP). Among the 16 exercises in the LAP, children must copy the letters y and h, copy a picture of a cat and a square, add seven parts to a person, cut out a diamond, count 10 cubes and tell how many there are, skip on alternate feet, and high jump 10 inches. Teachers interpret the results of this test in light of their own observations and informal assessments, but together these indicators of school readiness determine which children will go to regular half-day kindergartens and which will go to the extended day kindergarten. According to school policy, the more mature children will get extended day.

Only one other structural characteristic of Jackson related to testing. In five classrooms, teachers elected to use the Team Assisted Instruction program, a mathematics curriculum developed by Robert Slavin at Johns Hopkins University. This program requires that instructional groups of deliberately mixed ability be formed based on a placement test of computational skills. Otherwise ability grouping, either within or between classrooms, is absent from Jackson.

Compared with Jackson, Hamilton's organizational structure is more hierarchical, comprising a greater number of slices and layers. Students find their way into the categorical programs such as Project Potential, LD, and ESL in much the same way, as both schools follow state and federal guidelines in defining who is eligible for them. Yet what happens to children so identified is quite different. The ESL pupils experience a typical pull-out program where they leave their regular classes and come to the ESL classroom for special instruction for one, two, or three hours each day, until by test results and ESL teachers' judgment, they have sufficient grasp of English to get along in a regular program. The staff places LD students in transition classes along with children whose basic skills of reading, language, or math are more than a year below those of their peers. Teachers and administrators make placements to the transition classes following TAP meetings where they review test data, daily work, and teachers' observations. The principals use money the district allocates to Chapter One, the "Skills" program for primary pupils, and special education funds to operate the transition program. Such use keeps class size low throughout the school, averaging around 20 in regular and 15 in transition classes. Another function of this organizational structure is to keep children who might elsewhere be retained in grade with their age-mates in transition classes. The curriculum in transition classes is exactly the same as in the rest of the school. That is, while regular fourth grades are in Levels III, IV, and V of Reading Mastery and Spelling Mastery, the fourth-grade transition class might be working in Levels II and III. Most of the transition teachers have certificates in special education.

At Hamilton, the teaching staff uses test results to organize kindergartens. Teachers administer the Learning Accomplishment Profile and the Distar Language Test to incoming pupils. They will assign those with the lowest scores to the extended day kindergarten (the opposite arrangement from Jackson's). During the first three weeks of school, teachers make adjustments (into or out of extended day kindergarten) if children show levels of class performance or readiness different from what the tests show. Even though they rely on their own observations and judgments, teachers often use test results to justify placements, particularly to the lowest ability group. Rarely do parents challenge placements, however. Rarer still are upward moves, once pupils are in a particular level, even if their rate of achievement accelerates.

Teachers also use test scores to group children within classes. In the regular, half-day kindergarten, teachers group children into either a Distar group that works primarily on language development or into a Reading Mastery group that works on beginning sounds and letter recognition. These programs offer regular progression through a hierarchy of skills that lead to the first-grade program of phonics and further language development.

For grades one through six at Hamilton, tests play no role in allocating pupils to homerooms. Teachers in a given grade level meet and discuss which children might do best with which teachers at the next grade level. Principals adjust these suggested rosters to even out class size and make sure that no teacher gets more than his or her share of the particularly troublesome pupils. As already described, the exceptions to this procedure are the transition classes. There are very few parent requests for preferred teachers.

Although homeroom assignments are not based on ability or achievement, there is a salient hierarchical pattern for assigning pupils to reading groups. Dr. Michael groups children into reading "Levels" of Reading Mastery according to their previous year's progress in Reading Mastery (where they left off in the program the previous May). If that information is not available for a given child, he administers a placement test that is part of the Reading Mastery program. He then assigns teachers to Levels, some getting more than one. Even within each Level, some teachers make up reading groups wherein children progress through Reading Mastery at different rates. The following excerpt shows how Dr. Michael assigns pupils to Levels.

Grade Level Meeting at Hamilton

> Dr. Michael presides over the meeting of 10 Level V and VI teachers. He reminds them that a level in Reading Mastery is equivalent to where an average pupil in a grade level would perform. Level V, in other words, is where the average fifth-grade pupil would be placed. He has a green card for each pupil that contains ITBS and BST scores and progress reports from last year: the number of the last lesson in Reading Mastery that the pupil

successfully completed. He tells them that he "backed up 10 lessons from where they ended last year." Then he grouped "kids together who are within a 10 to 15 lesson span of each other and assigned them to a teacher. We hope to even out the load, so that each one of you has about 22. This results in three level V classes, one Level IV class and one transition class" for the fifth grade. Sixth grade has the transition class, two Level VI groups, and one Level V group. As soon as classes start, teachers will have to give placement tests to any new student without a green card.

In answer to a teacher's question, Dr. Michael explains that ESL pupils will have their reading instruction in ESL class rather than in a reading level. He goes on to explain the record keeping system they each must use. This records the number of Reading Mastery lessons completed in a ten-day period. "Ideally the average ability group should be doing about one lesson a day." The other data the teachers have to collect and turn in is the record of performance for each pupil and aggregated for the group; for example, the average percent correct on daily work and other indicators. Dr. Michael indicates that the teachers must grade according to a standard system, with teachers determining only about 10 percent of the grade, the rest the accumulation of daily performance and test results on Reading Mastery tests.

The teachers present seem content to follow Dr. Michael's system—at least there is no dissent. There is some puzzlement over a few pupils who, if the system were followed exactly, would be two levels above their grade. Someone suggests that they should be sent to the library, someone else jokes that these pupils teach should get a reading group of their own to teach. Another sixth-grade girl has been assigned to Level V even though her ITBS scores are above grade level. One teacher says she would like to see her bumped up, but Dr. Michael holds to his original placement, justifying it on the basis of the much greater "complexity" of Level VI compared to Level V. When Dr. Michael leaves the room, a fifth-grade teacher jokingly asks his colleague, "Do you remember that girl I had last year, whats-her-name? She would have to have been carrying a turnip to register an IQ."

In these ways, Hamilton organizes reading instruction. Teachers' discretion determines grouping for other subjects. For example, the sixth-grade teachers determine that they wish to organize by subject matter, with one particularly able teacher handling all the science, another specializing in language, and the third in social studies. Thus, at specified times in the day, pupils leave one set of classrooms and go to others for these subjects. In the third grade, the teachers have decided that they want to group by ability in math. The three teachers have either high, medium, or low-achieving pupils as determined by ITBS math scores from second grade. Third-grade pupils also exchange teachers and classrooms for separate subjects, such as science and health. First and second-grade teachers have decided that pupils will stay with their homeroom teachers for everything but reading. Two of the the fifth-grade teachers have divided their pupils according to the results of a placement test in math, with one responsible for lower-scoring pupils and the other teaching the higher performers. In all cases of between-class grouping by ability or subject matter, timing is crucial. Everyone must study reading on schedule. Otherwise the shuffling of children between homerooms and Reading Levels would be chaotic. Likewise, any other regrouping requires that teachers follow a rigid schedule for those subjects organized by department or ability group.

Although the OPENING AND ORGANIZING SCHOOL PHASE of the natural history of testing occurs primarily during the time before instruction begins, schools are reorganized periodically during the year. The role tests play remains the same; the staff uses test results along with other methods to reallocate learning opportunities, classes, and teachers to pupils. One sees this role best through observing the TAP meetings at both schools.

Hamilton TAPs

> It is 7:00 on a mid-September morning, and the campus of Hamilton is already humming as many of the pupils eat free breakfast. In the principal's office, the usual group assembles for the TAP meeting: Dr. Michael presiding, the school psychologist acting as official recorder and gatekeeper for any special education testing that might ensue an informal facilitator, the social worker, nurse, and, in turn, those teachers who are somehow involved with each pupil who has been referred for TAP. These meetings occur once in a six-day schedule if pupils have been referred since the last meeting. Business is so brisk today that there is a double-session, before and after school. The purpose of TAP, which exists throughout the district, is to provide expert consultation for teachers who are having difficulty with particular pupils, as alternatives to or prior to a referral for special education, for example. At Hamilton, according to Dr. Michael, "We tap everybody who changes placement, for the protection of the kids, and even if the school psychologist has not administered those tests. We look at the data available."
>
> Dr. Michael gets them right to work by calling for the first name on the list. Ms. Bennett has brought up T, who has run away again and hasn't been in her class for the last week. Ms. Bennett is not sure what to do with her if and when she comes back and wants her considered for a special placement, perhaps in the sixth-grade transition class. "If she comes back, will she come back into my class?"
>
> Looking at T's green card, Dr. Michael responds, "I assume so, she's not low on her skills," at least to the degree one would expect of someone needing special academic placement. Her last April ITBS grade equivalent score in math is grade equivalent 4.6, language is 5.6, spelling is 6.2; she is reading at Level VI. Dr. Michael talks about the horrible home situation and possibly abusive father. Bennett calls her a troublemaker in class. Dr. Michael says she maintained grade level work last year. The nurse suggests she should get counseling. Bennett asks if it is possible to get her classified as Educationally Handicapped so that she can get the help she needs. They discuss social welfare agencies. Bennett repeats her suggestion for the transition room. "It doesn't seem appropriate academically," says Dr. Michael, and Bennett agrees with slight resignation in her voice. "We'll have to use ATF with some judgment, because we can't have her facing the wall the entire time she's in class." The psychologist, who often comments on birth order, age in relation to grade, prior official diagnosis, and genetic causality, asks whether T is an only child. They exchange additional anecdotes about T's bad home life. Dr. Michael says the school can file truancy papers with the police, so that she will qualify for counseling through the social agencies rather than the schools.

By completing the paper work he must file on each child and calling the name of the next pupil on the agenda, the psychologist signals that this will be the extent of T's TAP.

The second pupil to be "tapped" is N, a pupil in Mrs. Marshall's third grade. Because third graders change teachers for most of their subjects, all the third-grade teachers are present to provide information. Marshall says that N is "in the low reading class and the low math. She's moved around a lot and I don't know much about her background. I guess my whole problem with her is that everything is just too much above-level for her. She's just so low. Her reading, she's also so slow. She is having trouble passing the check-outs." Dr. Michael examines the paper and notes that N's performance in reading is close to passing, but slow. Mrs. Marshall pronounces N, "Very low in language skills. I guess that is what concerns me most in any of the kids I bring up for TAP. She can't write a complete sentence. Most of the kids are doing okay. She can't understand basic directions, and she needs a lot of individual attention. If she gets individual help on following each direction then she does all right." Dr. Michael questions her on N's spelling and comprehension, and Marshall admits (in a very quiet voice) that N is doing all right, but that "with both of the kids I'm bringing up right now, I just don't want to let it get too far along, that's why I brought them up early. She just is getting to be very frustrated that she is not getting even this very basic stuff that we're doing. And right now it's second grade level as far as language. And I can sit down with her and go through the directions and go through a sentence and she still doesn't understand it."

Dr. Michael asks what her scores on language are and Marshall says, "Not severe, but not where she should be either." She admits that N's oral language is not deficient. Dr. Michael takes over and says that "at least so far, N does not seem to be too low, no lower than several others and not even low enough to ask for formal evaluation. Marshall says she doesn't know what to do to accommodate her in her class: "What should I do, put her in a group of three? I don't know how to work that." He says he will talk with her later about that. "It doesn't sound like there is that large a discrepancy that we should put her through the psychological evaluation," says Dr. Michael. He is referring to the requirement that a child's measured achievement must be significantly below intellectual ability to be considered as LD. There is some discussion about whether N was classified as LD in her former school, and the psychologist volunteers to call. Mrs. Marshall asks whether a child who moved in already so classified would automatically qualify for the transition class. Dr. Michael says in that even they would not even have to go through TAP to place the child in transition class. He asks Mrs. Marshall whether that was what she was seeking, and she says yes, otherwise she would have to form two language groups. N's math teacher seems to go over to Dr. Michael's position, saying "She really is doing okay with me." They all agree to the status quo.

They next consider M, who is also in Mrs. Marshall's homeroom, but because of his limited English, he goes to ESL for about an hour and a half each day, during reading period. The ESL teacher presents the case that his social language is adequate, but "Academically, it's just not there. So I thought maybe if he was in transition, he would have more one-on-one and they wouldn't go so fast." Dr. Michael points out that unless there is evidence for a dysfunction of some sort, transition class would not be the right placement. If the problem is that he doesn't have the language, then it's a different problem. The psychologist says that the transition program is for kids with real disabilities, and instead it is filling up so fast that the kids who really need it aren't being adequately served. The ESL teacher says that no one could have tested to see if he has real learning disabilities because he doesn't have enough English to be tested and there are no Iranian test givers.

Dr. Michael: "If we start moving kids to transition from all over the school, even those who don't have English-speaking skills, we'd fill up the classes fast." The ESL teacher continues pressing the case for a slower paced instructional program for just one year (the rationale being that it takes him time to translate instructions, etc., from Farsi to English, and that this constitutes a language-learning disability). The psychologist asks why the teacher can't "modify her program to adjust to his situation. There must be other children who have slower processing." A number of children are truly LD, and they need this transition. "I'm not saying he doesn't need help." Some suggestions are made about peer tutors and volunteers. The ESL teacher is asked to add him to her load, and she says her roster is full. Mrs. Marshall asks, "Isn't there something she can give me to help him; he has to have some help." Dr. Michael asks what the focus is, language development? Does he need more peer tutoring, more coaching, more concept work? Mrs. Marshall answers that all of that is needed and argues again that she could spend hours with him just explaining the directions on the worksheet.

Dr. Michael tries to summarize and bring the case to conclusion, asserting that there is no evidence for M's having a cognitive deficit that would qualify him for transition class. "We don't have any evidence that it is anything but language." Transition is not an appropriate placement, and a combination of time with ESL, tutoring by volunteers before and after school, and the teachers' adjustment of the classroom program will have to suffice for the time. Dr. Thorne, who has been listening but so far not participating, asks what the ideal placement would be, and there is general agreement that a full-time self-contained ESL class would be it. Since there are so many similar problems at Hamilton, the neighborhood having been designated an immigration site, Dr. Thorne will talk to people at the district office about getting more ESL resources.

The fourth case is a boy who has already been "tapped" last week. There is a fast two minutes devoted to alphabet soup: OLDC, IEP, EMH, DDK. The psychologist talks about the flow of paper work; they quickly make the decision to place the boy in another school that has special, self-contained programs for the more severely handicapped. They define the fifth case as a similar problem and deal with him the same way.

The sixth case is P, a fourth-grade girl. Mrs. Grady hands them a spelling test and notes that P's previous school had recommended her for LD. She has already repeated a grade. "She is a neat kid. But she has a lot of trouble with reversals, and I'm wondering if she is dyslexic. She does well on oral answers." Dr. Michael says, "I don't know about dyslexia, she might be dysgraphic" (the handwriting is so bad). Dr. Michael suggests calling the previous school to get diagnostic classification and testing records. Dr. Thorne comments that she is a little old to be trained in fine motor skills, but OT should make an assessment. If she does not have a testing record in the previous district, Dr. Michael says to bring her up for TAP next week. "I question whether she is appropriately placed in your classroom, but that fourth-grade transition is bursting at the seams. Our goal is to keep the pupil-teacher ratio down to 15 and she is up to 18 or 19 and that's with a full-time aid. But yet when I look at that kind of writing, and she's in Level III reading...." Mrs. Grady reveals that P is not in the lowest math group, but still low: "She does have some multiplication facts, and she is willing to try once she learns..." Dr. Michael takes over and says he will make the calls and then will consider moving her over to transition [apparently, he is convinced that P is low enough to indicate learning problems whether prior testing was done]. There is a general discussion about eye testing, OT testing, perceptual problems, and setting up a formal psychological evaluation. The transition teacher seems willing to add a pupil to her load. Dr. Michael asks whether P's oral expressive language is okay and Mrs. Grady gives evidence from her observation of P's class work rather than any test results. "The problem is just that when it comes down to getting things down on paper, her writing and spelling are so bad that I have to ask her to translate it." The decision to transfer P into the fourth-grade transition class is made on the basis of the teacher's judgment, Dr. Michael's concurrence with the evidence, and the below grade-level ITBS scores from the year before. Mrs. Grady seems pleased with the results of TAP.

One more referral, another of Ms. Bennett's. Although the ITBS scores of one of her pupils are about at grade level, he is "my anchor man" in all subjects and is struggling and failing in class work. She has seen his records from a previous school where he had been diagnosed as LD. "If he is LD, I want to know it, because he can be placed in transition."

So ends the TAP—much business for 45 minutes. Teachers seem to use TAP to accomplish a change in levels or class placements, primarily to transition classes. They use test scores as evidence to justify placements that they make and justify to themselves on other grounds. Dr. Michael and Dr. Thorne use test scores defensively against teachers' overuse of the transition program, which the administrators wish to remain solely for pupils who are LD and those who are more than one year below grade level on test scores and progress through Reading Mastery. Diagnostic testing, the province of the psychologist and the special education teachers, seems to be accepted as valid and appropriate, yet not always final or critical in making decisions about organizing classes.

TAP at Jackson follows different patterns and occurs less frequently, the first meeting not scheduled until October. The meetings are in the afternoons, with Mrs. Mitchell, the psychologist, and the social worker (the same ones who serve Hamilton), the AIM teacher (for children who have been staffed as mildly emotionally disturbed), the nurse, LD teacher, and referring teachers. The psychologist acts as recorder, keeping track of the flow of special testing and decision-making, and Mrs. Mitchell presides. The following TAP session at Jackson took place in early December.

The psychologist, eager to get things moving, raps on Mrs. Mitchell's closed door and urges the others to convene. Mrs. Stevens has referred two of her third-grade pupils. M, as Stevens describes her, has a "reading level that is very low. She has high distractibility and never stays on task. She is jumpy all the time. Her IQ is okay, and she is showing that she can hold her own in math, so why isn't she reading? Her English is very poor, her grammar and tenses."

The psychologist says, "When you mentioned language, I wonder whether she comes from a jumpy family and she may be a candidate for ritalin." Mrs. Stevens is incredulous. The social worker, apparently thinking that M's problem may be environmental rather than physiological, asks whether M is less jumpy in art class. The AIM teacher chimes in with a possible need on M's part for relaxation training. Or possibly these symptoms indicate that her family may be on drugs. The psychologist mentions that ritalin is very effective in treating Attention Deficit Disorder and recommends a trial dosage to anyone with a particular pattern of scores on the Weschler Intelligence Scale for Children, which he will administer to M. "If the trial dose doesn't work, then the problem is something else." The social worker again recommends observation of M in non-academic settings and cautions that problems of behavioral origin are not affected by ritalin and that many teachers are oversensitive to this problem and likely to attribute conduct problems to physiology.

Mrs. Stevens tries to bring the discussion back to academics: "You know what's strange? In kindergarten she was graded satisfactory in everything, but in first grade she was marked unsatisfactory." The psychologist attributes this change to the kindergarten teacher: "Maybe she didn't care enough to grade." Mrs. Mitchell retorts, "Maybe the first-grade teacher screwed her up." Mrs. Stevens again: "My concern is reading." The psychologist refers to M's ITBS scores of 1.6 in reading in second grade. In spite of the social worker's cautions and Mrs. Stevens' questions, the prevailing view is that M is ADD, and this is the direction that subsequent testing will take.

Mrs. Stevens' other referral is A, who is "very, very low in reading. On the first grade ITBS, she got a 1.0 and on the second grade ITBS she got a 1.3." Someone sarcastically notes that the scores represent improvement. Mrs. Stevens says that A has a good IQ and good math background, and "I want to know what to do to get her reading up. I want to get some good testing done so that I can know what channels to use with her." The Chapter One teacher has come in and offers to repeat a Reading Miscue Analysis with A. She says that A can get a story from context, but seems to have regressed a great deal over the summer. Mrs. Mitchell says she will listen to A read, as a kind of informal assessment of the kind of reading in context and for purpose that is so important to Jackson. This is the resolution of the academic problem, but Mrs. Stevens and Mrs. Mitchell discuss how the mother tries to hustle them, can never get A to school on time, has moved around so much and had so many marriages and boyfriends and general instability. Abuse and instability are common themes in TAP.

The next referral is by Mrs. Norton, the teacher of the transition first-grade class. She is here because of an automatic referral mechanism built into the district TAP guidelines. Pupils whose performance is below certain cutoffs must be tapped. "The only reason I'm here is because someone says I need to be. L is making progress. She's properly placed. This is her fourth school in two years. Testing was never done because she always moved just before the testing was about to be done. She is in developmental first. She looks young. She's small. I think she is LD, but, hey, I've seen progress. She talks a lot now. She does know her alphabet and she knows all her sounds. Her art work is superior. She likes books, she likes to handle them and listen to them. She has problems counting with one-to-one correspondence. She hasn't had as much experience with all the materials like the other kids. I think we need to touch base on her; we'll eventually have to test her, but we don't need to do anything else now. The speech teacher will be seeing her; she's already been screened." Mrs. Mitchell interjects a few questions and thanks Mrs. Norton for her contribution.

This is the final tap for the day, but there are more comments on the side. The social worker asks Mrs. Mitchell if she can begin a group for children of alcoholics and whether there is any chance of running a parenting class here. The psychologist and the AIM teacher share war stories about particularly troubling children and their bad home lives. The psychologist suggests that a child who has been giving the school trouble be taken for a ride to see the nearby reform school and be told that that is his destination if he doesn't "choose" to shape up. The LD teacher glares at him, but he takes no notice.

Because there are fewer "levels" and ability groups at Jackson than at Hamilton, the focus of TAP is less likely to be on a pupil's measured ability and the outcome less likely a change of assignment. Test scores come less often into play. Yet those reallocations that take place are frequently just as severe; for example, changing pupil identification from "normal" to LD or emotionally disturbed, or transferring pupils out of school into self-contained programs for the handicapped.

Testing at Stage Two. The role testing plays in this early stage in the cycle is to set boundaries around the possible learning opportunities or the face-to-face groupings available to pupils. These help define for the pupil what he is and what he can possibly become. The fifth grader who reads well enough to be placed in Level VI now has sixth graders as immediate reference groups. Children in the Jackson transition first grade now are separated, perhaps for their entire pupil career, from their age mates in regular first and perhaps have lower expectations and fewer chances to learn. The child whose teacher judges her to be in need of remediation is denied the transition class because her ITBS scores are too high. Although secondary in importance to other indicators, such as last year's progress through Reading Mastery (to Dr. Michael) or oral reading of a story (to Mrs. Mitchell), test scores are a kind of arbiter of school structuring decisions. Overruling the arbiter takes strong arguments and courageous actions by educators.

Stage Three: Putting Tests in the Background, Ordinary Instruction in the Foreground

By early September, children and teachers have their niche, last year's test results fade from teachers' memories, and the spring tests are too far in the future to pose much threat. Now the teachers direct their attention toward the attainment of educational goals the institution or they themselves set. Only by close examination of classroom life can one see the relation of external tests to teachers' actions in the course of ordinary instruction. As findings in Chapter Two make clear, teachers frame many of these goals not as outcomes but as processes such as building a sense of community and instilling love of reading. Tests, either internal or external, help little if at all in judging the success or failure of such ventures. When teachers define educational value and attainment as outcomes of instruction, it is possible to ask what role tests play.

In this section we consider the means by which teachers assess pupils' progress toward valued outcomes and categorize them as follows: autonomous teacher judgments, results of external tests, curriculum-embedded tests, teacher-constructed tests, and CUES (Continuous Uniform Evaluation System). By presenting vignettes of classroom life, we attempt to show how such assessment takes place in ordinary instruction and how it relates to external testing.

Autonomous teacher judgments. Certain assessments in ordinary instruction at both schools rely on teacher observation and judgment rather than formal tests. Teachers using Math Their Way as it was developed—with authentic problems in numeracy (counting real objects in the environment, tracking the calendar, counting and graphing real events)—have the children using concrete objects such as unifix cubes to understand concepts. They rely on their observation of how children use the materials to judge whether the concepts have been grasped. Teachers using SCIS at both schools trust observation in the context of problems and lessons to judge the children's grasp of concepts. Teachers at Hamilton who elect to use the district's program *Writing and Thinking* rather than *Expressive Writing* must assess and grade by judging pupils' effort, progress, interest, and mechanics.

In most other respects the two schools differ, with teachers at Jackson using their own experience and sense to assess pupil progress. This takes several forms. In reading, the primary mode of instruction at Jackson is the literature study. The class, interest group, or individual selects a book (literature rather than a textbook). The teacher discusses with the children, or listens to the children discuss among themselves, the features of the book: characters, setting, story line, and illustrations, and compares it to other books by the same or other authors. The teacher judges the pupil's progress by the extent of interest and participation, whether the features of the book have been noticed, concepts grasped, and vocabulary understood or questioned. Assessment and instruction interweave as the teacher suggests resources the pupil might pursue to find out the spelling or meaning of words or discover background information on a particular subject, all of which the teacher monitors. Writing and reading are integrated, as teachers frequently ask pupils to draw pictures of and write about parts of the stories they read. Again, the teacher is checking for effort, involvement, participation, and progress toward understanding the author's intentions and the meanings of words within the context of the literature. Teachers frequently talk about how pupils handle the materials in class, how they listen and ask questions, by what process and criteria they choose books to read, how they attack unknown words by decoding strategies and context cues, and their fluency in reading. When the teachers discuss assessment, they demand that it be "in context"; knowledge and achievement must be considered in light of specific lessons, books children read, and projects they collectively brought to completion. Because of the absence of what one would normally think of as standardized hierarchies of materials to be covered and skills to be learned, assessment for the purposes of accountability, communication with parents, and the assignment of grades trouble teachers. The solution for one of the primary teachers is to use rating forms reproduced in Exhibit One.

In Mrs. Shaw's Rooms

> Without fail, every morning at 11:00 Mrs. Shaw summons her second graders to the area rug in front of her maple rocking chair, where she reads them a story. Her intonation rising and falling, the tempo changing with the story line and dialogue, she allows them the sheer pleasure of hearing words and evoking meaning. She reads the book straight through, not showing the illustrations or asking what they predict will happen next. By the second week of school the children already know that comments or questions must wait to the end. Then come the conversations and exegesis: "What other stories does this remind you of?" "Let's read the dedication." "Look where this book was published." When she asks them questions, they are open-ended ones in which she clearly indicates she is interested in what they are thinking. "Well, what do you think of Chris' new dog?" The questions are not asking for details they remember from the story or literal comprehension such as one sees on reading achievement tests (e.g., "What was the name of Chris' new dog?" or "Which of the following would be the best title for this story?").
>
> The classroom is awash with text: reference books everywhere, books for literature study or for story time, separate stacks of literature and personal experience journals; stories or papers the students have completed are tacked up on the wall, taped to the chalkboard, or hanging from a wire strung across the room overhead; a list of favorite words they have encountered in stories, a message on the computer screen; charts and graphs from Math Their Way where children are keeping track of the number of games the Phoenix Suns win and lose, the number of days the weather is sunny or stormy, pictographs representing the attendance of students in the class.
>
> By the time Mrs. Shaw gets to story time, the children have a day's work behind them. They have done the calendar activities in Math Their Way and have gone through an activity on place value using beans and medicine cups. While some children are in the writing center, others will be reading and writing in their literature logs or working on a science activity. Mrs. Shaw is an avid user of SCIS, a hands-on science program in which children conduct experiments, make predictions, observe the results, draw conclusions, and discuss and write about what happened. Right now, they are working on concepts of mass and volume with a balance scale. Although many teachers have given up on SCIS, saying it takes too much time to set up the experiments and order the materials, it is too messy and pupils are too noisy when they do the activities, Mrs. Shaw and some others feel that "this

is the only way children can learn science, certainly not out of a textbook." She feels that doing science gives the children something to think and write about, that science makes cognitive processes available to children, that schooling has to be more than just an acquisition of skills.

The children in Mrs. Shaw's class are divided into five "writing process" groups. She meets with each group once a week for 50 minutes. The rest of the class works independently on projects she has structured, sometimes with assistance by a teacher's aide. The writing process group meeting on one typical day in March consists of Ben, Ashley, Elizabeth, Darby, and Jon. Except for Ashley, these children are typical of the rest of the class. Mrs. Shaw often uses Ashley's writing style and language structure as mini-lessons for the group or the class as a whole. "You can always tell what we are learning in our classroom by looking at Ashley's stories," which now include mystery stories about Egypt. Reading a book suggested by Mrs. Shaw prompted this interest and led them to a whole investigation about ancient Egypt, mummies, and hieroglyphics.

Mrs. Shaw gathers the group in a circle on the floor and asks who wants to begin. Ben wants to read his story about Freddy, a character in a horror movie whose frequent mention by school children causes their teachers to cringe. In his own words, Ben has written a story that recapitulates this movie, but Mrs. Shaw cuts him off. "Ben, you've been in this class all year and last year you were in this school, so I know you know that stories like these are not going to be allowed in our Young Author's Day celebration. They are fine for writing for practice, but it's not a story we allow to be published as good literature. We think you should be writing good books by this time."

Ben: "Yeah, I know! I wasn't going to publish it. I want to publish my whale book."

Mrs. Shaw: "I agree. That's an interesting story! Who wants to be next?"

Ashley reads her story, *The Mystery of the Stolen Mummy,* fluently and mysteriously, and the children lean into the circle to hear every word and look at the written words as she reads.

The Mystery of the Stolen Mummy

I went down Main Street to the corner. I heard footsteps behind me. I looked behind me. No one was there. "Maybe it is just the wind," I said. I kept walking. I turned the corner. When I turned the corner, I saw an old, old house. so I went in the house.
I saw a mummy case in the corner. So I walked over to the mummy case. I opened the case. Nothing was there. But there was a note. It said, "Beware!" Beware. Beware. What does it mean? Then a voice said, "Beware!" I slowly turned around. A man dressed all in black was looking at me. Then I asked, "What does Beware mean?" But he just laughed and walked away. Then I fell through the floor. It was a secret passageway. I saw a door. I went in the door. I was outside. There was no door. I was confused! I tried to remember everything, but I couldn't. I remembered the mummy case. There was a symbol on it."

The children waited for the end of the story, but Ashley had not finished it. Rather than asking her how the story would turn out, Mrs. Shaw focuses their attention on the kinds of information and detail Ashley might add to make the story more credible, a better story.

Mrs. Shaw: "Oh, that's exciting! Isn't it?"

Elizabeth: "I helped her. My story is about mummies and trap doors too."

Mrs. Shaw: "I can tell you guys have been reading a lot of mystery stories. Are the symbols in your pictures going to match the symbols on the mummy? Maybe they could be hieroglyphics? Do you know what hieroglyphics are? That's Egyptian writing. Do you see that book back there, *Alphabet in Ancient Egypt*? You could use that book and find some alphabet letters written in symbols. That could be part of your pictures when you illustrate your story. Go back and get that book."

Elizabeth: "I know Indian symbols."

Mrs. Shaw, turning the pages of the book: "See, the Egyptians did their writing and they did it in a language called hieroglyphics."

Jon: "I was looking at this book before and I noticed that."

Mrs. Shaw: "Let's look in the Index under H. Here it is, page 18. Here are some. You've heard of King Tut? Well, this is how he wrote his name."

Children: "Weird!"

Mrs. Shaw: "It says here that the set of symbols on the right are the two T's. That must be the U (indicating the hieroglyphic of the bird). You have this page, so you could figure it out [include a hieroglyphic symbol in Ashley's story]. That would be a neat thing to do, wouldn't it? You can keep this book for a while. You've got a good story going there. I hope you get it finished."

Elizabeth reads her mystery story that takes place among the pyramids. The group makes one suggestion about describing the characters in greater detail. Now she is ready to "publish" her story, a process that involves editing it to find and remove errors of spelling, grammar, punctuation, and capitalization (first by other students, then by the teacher) and then entering the story into the computer, printing it out (the teacher does much of this), and illustrating it.

Now it is Darby's turn to read.

Darby: "I don't know if this is going to be too good a story."

Mrs. Shaw: "I like the title. What's the problem?"

Darby: "Well, these people that rob a TV set."

Mrs. Shaw: "Do you just want to read it to us? Do you know the problem in the story right now? The robbery?"

Darby, reading the story: "Jill was 16 years old. She had a big sister. Her name was Chrissy. Jill had a twin sister. Her name was Joanne. Last night we got robbed. The robbers—"

Mrs. Shaw: "Now, wait a minute. Stop! Are you in this story too? Don't you mean to say, 'They got robbed?' They got robbed."

Darby: "Yeah. The robbers stole our TV and our VCR."

Mrs. Shaw: "So it's their TV and their VCR."

Darby: "Yeah. My grandma is kinda rich and they got stolen a lot, like three times."

Mrs. Shaw: "So you could talk about things that happened to them to put in your story. What really happened? That would be good."

Again Mrs. Shaw directs the students to facts and information that give credibility to stories. She mentions references, witnesses, and other sources of information they might tap. She leads Darby through a series of questions much like a crime investigator or lawyer would, to extract details of the real incident from her: the time, setting, first thing that happened, second thing, people's reactions, how they felt, what they said, what the police did. She sums up this activity of reportage by saying, "You'll have to make up your own ending. If you're going to catch them or not. But that gives you facts about how the police act and do everything. Maybe then the girls could go off and follow the footprints after all this happens in their house, and then they could solve the problem in some unique way." At this point Mrs. Shaw sees that the hour is up and calls the class together for the story.

For Mrs. Shaw, assessment serves instruction and is intimately tied to it. After a period of time getting to know the children, she observes them working. She carefully records how they select materials and use them, and how they interact with others. If she gives them a worksheet, she is more interested in the processes the children use than the skills that the worksheet is supposed to teach. She notes how children work collectively on certain tasks, believing that learning is a group accomplishment as much as it is an individual one. She keeps many records. For example, she has the children read a story into a tape recorder in the fall and spring of the year so that parents and the children themselves will see the progress. She does the same with early and late samples of their writing, and a permanent folder of each child's writing is available. Working with the other primary teachers at Jackson, Mrs. Shaw has developed checklists for evaluating pupils' accomplishments in math and language, reading and writing. On the literature study assessment, the criteria she checks off include: "completes book reading according to group plan," "participates in discussion by recalling other books which build relationships and connections," and "participates in discussion to support own ideas or discovers new ideas from the group." Like

> many Whole Language teachers, Mrs. Shaw watches carefully for how her pupils' "skills" emerge. In an environment rich in language and text, she believes, one does not need to teach such things as phonics and word attack skills separately. Instead, she teaches such skills when children need them to comprehend a story or write for real audiences. One can see how standardized testing, which defines achievement as acquisition of common skills and correct answers to closed-ended questions, would be antithetical to Mrs. Shaw's brand of teaching. In her class, not everyone learns the same things at the same times. A different writing process group might not consider Egyptian topics at all and as a consequence would be learning different sets of background facts.
>
> When Mrs. Mitchell evaluates and supervises Mrs. Shaw, she observes, takes detailed notes, and sometimes videotapes. Later, Mrs. Mitchell and Mrs. Shaw will read the notes or watch the tapes together. As she does this kind of supervision with each teacher, Mrs. Mitchell encourages them to reflect on what happened and whether what happened was what they intended.

Results of external tests. If one restricts the definition of external tests to the ITBS and BST, it is safe to say that no teacher uses the results of external tests in the process of assessing the progress of pupils toward goals or the assignment of grades. Testing and assessment are perceived as separate activities that never intersect, a finding that observations and interviews reveal. Illustrative is the following quotation from a intermediate grade teacher at Hamilton:

> I know what the kids are doing. I don't have to take the ITBS to tell. I know what they've learned. I know where they started, and I know where they are at the end of the year. I don't need a test to confirm what I already know...The results of the test won't affect their grades one bit.

By Arizona statute, schools must determine who passes third and eighth grades by cutoff scores on district tests, and we found some schools do make promotional decisions in this way. The two schools we studied, however, pay almost no attention to these scores in making the decision. Instead, the schools rely on teachers' own methods of grading to determine who passes. Teachers sometimes use the scores to lend credence to the decisions they make on other grounds.

Curriculum-embedded tests. At Hamilton, most assessments in ordinary instruction consist of tests embedded in the textbooks the building or district administrators select. In Reading Mastery, children of similar abilities proceed through a hierarchy of instructional materials ("Lessons" consist of adapted stories, vocabulary lists, worksheets, and drills—also phonics recitations in primary grades) and periodic "Checkouts" (individual oral readings of parts of the stories or word lists) and Unit tests. To satisfy the principal's expectations for normal progress through the materials, teachers must move groups through approximately one lesson per day and file paperwork biweekly to show this progress. Pupils must pass the checkouts with fewer than a stated number of errors to progress from lesson to lesson in a unit and pass the unit tests at given criterion levels to proceed to the next unit. If the group of students falls below the criterion number of errors in a lesson, they must, as a group, repeat the lesson, rereading the stories and lists, then retaking the assessment. If numbers of pupils fail to achieve the passing criterion in a unit, the teacher takes them back through the lessons. Sometimes groups break into two at

Exhibit One

2nd Reporting Period

__________________ student's name

MATH

counting by ones

1 = 1 - 100 __________

3 = 1 - 50 __________

counting by tens

1 = 10 - 100 __________

BL = can't __________

counting by fives

1 = 5 - 100 __________

2 = 5 - 80 __________

3 = unsure __________

counting by twos

1 = 2 - 30 __________

2 = 2 - 20 __________

3 = unsure __________

counting backwards

1 = 20 - 0 __________

2 = 15 - 0 __________

3 = 10 - 0 __________

addition and subraction

1 = understands and can demonstrate both combinations in addition and subtraction __________

2 = understands addition, some difficulty with subtraction __________

3 = unsure of addition and subtraction concepts __________

tubbing time

1 = extends activity __________

2 = loses interest after a time __________

3 = changes tubs often __________

AVERAGED GRADE __________________

Exhibit One

LITERATURE STUDY PLAN

NAME OF BOOK ______________________________ DATE OF STUDY ________

NAMES				
PREPARATION FOR LITERATURE STUDY				
1. Completes book reading and knows story.	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5
GROUP SHARING				
1. Talks about book.	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5
2. Reads selected parts to support ideas.	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5
3. Recalls other books which build relationships and connections.	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5
4. Is familiar with:				
Character	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5
Setting	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5
Tension	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5
Problem/plot	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5
Ending	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5

Exhibit One

LITERATURE STUDY PLAN

NAME OF BOOK ______________________ DATE OF STUDY ______________________

NAMES						
PREPARATION FOR LITERATURE STUDY						
1. Completes book reading according to group plan.	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5
2. Reads book for "just reading" and records initial reaction in lit log.	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5
3. Prepares for book study by rereading book and marks passages, writes comments or questions in lit log.	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5
LITERATURE STUDY						
1. Talks about the book from notes.	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5
2. Reads selected parts from the book, or refers to sections to support ideas.	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5
3. Participates in discussion to support own ideas or discovers new ideas from the group.	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5
4. Enhances discussion by recalling other books which build relationships and connections.	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5
5. Plans with the group for selected indepth study.	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5
SELECTED STUDY						
1. Prepares for discussion by making notes or projects to support study of Character Place/Setting Plot/Structure Time Symbol Mood T[illegible] on(Incident	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5

this point, one subgroup forging ahead and the other recycling through the same lessons as before. But there is no room for either acceleration or for individualization. The optimum group size is seven or eight in primary levels. If all but one member of a group needs to repeat, the one moves back with his or her group with no opportunity to move at a faster pace or move up to a more able group. When a group (or a whole class in the intermediate grades) completes all the units in a level, it proceeds to the next level.

Besides these formal assessments in Reading Mastery, pupils accumulate a vast array of scores on daily work, such as percent of answers correct on the daily "Skill Drill" worksheets. The principal at Hamilton requires teachers to use all these quantitative methods to arrive at grades in reading for the quarter. Some teachers add to or alter this requirement. For example, one intermediate grade teacher specifies that his pupils cannot earn a 1 (A) for the quarter, regardless of their accumulated test scores and daily work, if they have not completed a certain number of book reports. In general, however, most teachers adhere to the standard practice of assessment and grading that Reading Mastery curriculum specifies.

Although Reading Mastery at Hamilton seems to follow its own definitions of educational attainment without regard to external testing, a close look reveals concessions made to the ITBS. For example, the principals have required that primary grade teachers prepare seatwork packages for their pupils. Thus, while teachers are working with reading groups of six to eight, other children work at their desks on worksheets. The content of these worksheets fills the gaps between the methods of Reading Mastery and the contents of ITBS. Reading Mastery concentrates on phonics skills taught in idiosyncratic ways using altered text forms (special markings on letters and syllables that pupils are trained to associate with certain sounds). Worksheets use conventional forms so pupils will be accustomed to them. Comprehension of stories is not a part of Reading Mastery until Level III, so its developers recommend supplementing it with worksheets on "inferential comprehension." One worksheet SRA publishes for Level II is as follows: "There's a picture hidden on this page. To find the picture, answer the questions in the spaces. If an answer is yes, color the space pink. If an answer is no, color the space yellow." Among the questions are, "Can you ride a bike to the moon? Is grass red? Do sharks have teeth? Can a boy talk? Can a dime swim?" Another supplement is formal instruction in logic through the SRA program, *Thinking Basics*. The program teaches these thinking skills: classification, inference, deduction, definition, similarities, analogy, and description, and uses the teaching style of choral recitation repeated until the teacher detects no errors. For example, the teacher reads from the manual, "The first thinking operation is classification. Remember the rule for classification: If a class has more kinds of things, it is bigger. Everybody say the rule." The class responds by repeating the rule in unison. "The class of buildings has more kinds of things than the class of houses. So tell me which class is bigger." "The class of buildings" is the required response. If recitation is ragged or the teacher hears errors, she rereads the rule and the question until recitation is free of errors. "How do we know the class of buildings is greater than the class of houses?" The correct response is, "Because it has more kinds of things." There are several other examples the class works in this way. Then the teacher proceeds to the next thinking operation.

Assessment in mathematics follows the Macmillan textbook, for the most part. Scores on sets of problems provided by the series accumulate as daily work.

Tests are almost all from the text. Exceptions in primary grades are those teachers who have elected to use the program of manipulative materials entitled, "Math Their Way." There is a certain uneasiness among these teachers, who are philosophically committed to the use of concrete materials to instill understanding of math concepts. Yet they worry that they will fall short of expectations for normal progress through the text. The tests of achievement loom as distant harbingers of accountability for progress they will make. The form of external tests more closely resembles the materials in the mathematics texts than the manipulatives. As the year goes along, teachers scramble to include worksheets and problem sets along with the exercises in Math Their Way. In this way, teachers who stray from the norm, where the text defines instruction and assessment, attempt to satisfy their own educational goals (for conceptual understanding of math) and the institutional goals for training children to master math facts.

By looking at the text, one can see the influence of external testing. For example, the Macmillan series presents, across all grade levels, specific units of instruction on how to solve story problems. Even in primary grades, children learn that when they see the "clue words, all together," they should add the numbers. In intermediate grades, children learn the "steps in solving problems": read the problem; plan what to do (think about what operation the problem calls for); do the arithmetic; give the answer (in the proper metric); check the answer (backward arithmetic). Since problems outside of textbooks and tests rarely take this form, one possible explanation for stressing such instruction is to prepare pupils for external testing. Teachers, however, accept it as ordinary instruction, or even as imparting "survival skills," as Dr. Thorne characterizes them. Other concessions to the ITBS include speed drills in arithmetic computation, horizontally oriented arithmetic problems, and practice on the terminology the ITBS uses, such as "number sentence." These concessions are built into the text series so that teachers accept them as ordinary instruction.

At Hamilton, testing in other subjects—language arts, social studies, health, spelling, and science—parallels the assessments built into the texts and adopted programs. One could examine the chapter or unit tests in these programs and capture the bulk of testing in ordinary instruction. Teachers have some discretion about the extent and intensity of coverage of texts as well as the use of tests and weighting of test scores in assigning grades. For example, there is no school-wide standard procedure for determining the performance to be judged as passing or failing, nor for progressing through the text. Some teachers might regroup those pupils who fall behind the majority of the class; others might simply give them low grades. The series the district adopted include Harcourt Brace Jovanovich (HBJ) in spelling, language, social studies, health, and science, although the primary science program allows for the use of the SCIS materials. To teach writing, the district suggests use of *Writing and Thinking*. However, most of the teachers at Hamilton substitute the *Spelling Mastery* and *Expressive Writing* programs, both by Science Research Associates and both using the familiar teaching formats of Reading Mastery. Most of the text-embedded tests conform to achievement testing models of assessment.

Curriculum-embedded tests also characterize some subjects taught by some teachers at Jackson. Those teachers who let the texts define their programs in spelling, social studies, math, or science also use chapter and unit tests to assess progress and provide a basis for grading. With a greater degree of teacher discretion at Jackson, there is more variety in testing practice. For example, one intermediate grade teacher might be an expert in social studies and specialize in developing units of study on topics of geography or politics or history, departing from the textbook series entirely. That same teacher, with less expertise and interest in science, might rely on the science text to provide instructional material and tests. Assigning grades and charting progress toward school or teacher goals—even when the texts are the source of tests—fit the priorities of the individual teacher.

Teacher-constructed tests. In our observations of classrooms at the two schools, there were relatively few instances of assessment of pupil progress by tests written by teachers. An obvious exception is in the teaching of spelling at Jackson, where teachers construct spelling word lists from the writings of the pupils. For example, in the writing of journal entries, stories, or reports, pupils are encouraged to use words for which they know the meaning but not the spelling. Later, in conferences, the teachers circle misspelled words and some other mechanical errors and ask the pupil to look up and correct the words in an edited version of the paper. Some of these words would then be added to the spelling word list for later testing. Other words might be added to the list from instructional units. In general, however, teachers rely either on their own direct experience or curriculum-embedded tests for assessing pupil progress in ordinary instruction.

Continuous Uniform Evaluation System (CUES). Arizona statute requires that every school district conduct a CUES minimum competency testing program, but leaves the contents, schedule, format, and other details to the individual districts. In the Cactus District, teachers administer reading, language arts, and math CUES throughout the year on a schedule that they deem appropriate for their programs. However, they must report results three times per year (down from six times the previous year). Almost exclusively, the tests employ multiple-choice or fill-in-the-blank formats.

The intended purpose of CUES is to assess district curricula, with tests tailored to the district's Scope and Sequence. As such, CUES would be defined by testing experts as internal tests (Dorr-Bremme & Herman, 1986). Such characterization is consistent with the view of those district administrators most responsible for testing. According to their view, CUES are the way the district has of ensuring that the grade level skills specified in the scope and sequence are covered; the tests provide a framework, particularly for the new teachers; they provide a management system by which the teacher can group pupils for instruction and keep track of and recycle instruction for those pupils who have not mastered each skill; they are short-term, a formative assessment for which the Basic Skills Tests are summative. The district intends that the objectives are minimum; that teachers would build on the basic education they suggest. The district scores and reports CUES results by percentage correct, aiming for 100 percent mastery. The testing department reports results for teachers and by grades within schools.

If the administrators of the testing department notice that a teacher or grade level at one of the schools falls below the 90 percent mastery criterion, they alert the relevant curriculum coordinator, who then meets the staff at the school to discuss what has happened and what might be done. One district administrator explains the official role of CUES:

> I say that no matter what method [or program of instruction] you are using, no matter what material you ascribe to in Jackson or Hamilton, you need to be aware of what the CUES skills are for that grade level and teach accordingly. And that will help you to fill any gaps that may exist in a particular approach or a particular set of materials to make sure that you cover a grade level or a district or a CUES cycle.

According to this view, there should be no need to depart from ordinary instruction to administer CUES. Instead, the teacher teaches a skill in the curriculum, decides when the skill is mastered, and administers the relevant CUE. Those who pass it go forward in the curriculum, and teachers reteach the skill to those who fail until they attain mastery.

Few teachers agree with this characterization of CUES. More likely, teachers administer CUES on a pace more responsive to the required reporting dates than to the placement of the skill in a sensible instructional sequence. When time nears for the reporting of CUES, the teachers cease what they are doing, conduct a mini-lesson on the CUES skill, and administer the CUE, later recycling through the mini-lesson for those pupils who do not pass. In some cases, teachers administer all the CUES at one time, even before relevant instruction, so that they will only have to do mini-lessons for those particular CUES that students have not passed. They do this to minimize the extent to which CUES interfere with ordinary instruction. With few exceptions, teachers regard CUES as an unnecessary, bureaucratic intrusion on ordinary instruction and irrelevant to what they regard as their true mission. At Jackson, the intrusion is on Whole Language instruction and work on teacher-initiated units, which are quite outside the domain of CUES and, in most cases, the Scope and Sequence as well.

Ordinary instruction at Hamilton is the Reading Mastery program, as well as Direct Instruction in language and spelling, all of which fail to articulate with CUES. For example, the district's objectives involve reading comprehension in grade one, while Reading Mastery postpones comprehension and concentrates on phonics skills. At the beginning of the year, Dr. Thorne gave notice that, because Reading Mastery was so distinct from the district Scope and Sequence and basal series (Ginn), Hamilton would be exempt from reading CUES. At some later point, however, this exemption seemed to have been withdrawn without much explanation. In January, the grade level chairmen were given notice to administer and collect reading CUES along with those of other subjects.

District curriculum coordinators have tried to integrate CUES with chapter tests in the Macmillan series, so that teachers who follow the math textbook can administer CUES tests as they normally proceed through their math curriculum. But this fails to work out because the Scope and Sequence in math, and hence the CUES, require many recyclings through material such as using clock times or coins to compute addition and subtraction problems. The Scope and Sequence call for such skills to be introduced at one grade, mastered at another, and reviewed at a third, so CUES on these skills must be repeated. Where the text fails to cover material the Scope and Sequence specifies or sequences skills differently from the text, teachers must interrupt ordinary instruction to teach and test the CUES. For example, metric measurement is emphasized in the sixth-grade Scope and Sequence but is not well covered in the text. Furthermore, teachers who use math manipulatives to the exclusion of paper-and-pencil instructional methods are out of step with math CUES.

At the beginning of the year, Mrs. Mitchell announced that Jackson would be one of three elementary schools in the district that was participating in a pilot program of CUES testing in language arts. This program, the joint product of the district curriculum director for language arts and a committee of teachers, departs from standardized testing formats typical of other CUES. Instead, the teachers are to record on a machine-scorable answer sheet a rating for each pupil at each CUES reporting date. The ratings are to be based on teacher observation and judgment about whether or how well a pupil had satisfied each of the criteria for composition, oral communication, editing, usage, and language concepts (grammar). At grade level meetings early in the year, Jackson's teachers responded favorably to this pilot program, feeling that it was better suited to their Whole Language program than were the standardized tests of the official CUES. By mid-year, however, Jackson had abandoned the pilot assessment program and returned to the regular CUES in language arts. Mrs. Mitchell reported two reasons for this decision. First, the teachers felt that even this pilot form of CUES represented a major departure from ordinary instruction as they conceived of it. Second, they perceived that the regular CUES substantially overlapped the Basic Skills Test in grades three through six as well as the ITBS in all grades. Because of this similarity in format and content, they and their students were disadvantaged by redirecting their instructional and assessment efforts toward the pilot CUES. As long as they were going to be held accountable for the content and format of BST and ITBS, they might as well prepare their pupils by instructing for and taking the official CUES.

Among teachers and principals, CUES have few advocates. Some say that they are so easy that pupils can easily pass them, but high rates of mastery do not show up later when they take the BST. Others say that some of the items are poorly written and confuse the pupils so that they make errors just because of bad test construction.

Although district administrators perceive the CUES as an integral part of ordinary instruction and a means of controlling what is taught, teachers fail to share this perception, seeing them instead as an unjustified, bureaucratic intrusion on ordinary instruction.

To see CUES only as a device for controlling what is taught or as an unwarranted intrusion is to miss another, perhaps latent, function CUES play. CUES provide a structural link between this stage and later stages in the natural history of the testing event. By virtue of the similarity in item formats and content between CUES and ITBS, teachers are already, and perhaps unknowingly, preparing pupils for the external tests upcoming, and altering instruction accordingly.

The following vignette illustrates ordinary instruction and classroom life at Hamilton School. In addition, it shows how CUES testing is used and the roles assessment can play. Material for this vignette was pieced together from observations we made in several different intermediate grade classrooms. It appears here because it illustrates typical classroom life and supports assertions we make about the role of testing.

In Ms. Engle's Fourth Grade Class

The P.A. announcements barely interrupt children who have been intently writing in their journals for the last 10 minutes, while a cassette plays a tape of Madonna that one of the children has brought in. Since the *Expressive Writing* materials have finally arrived, time for journal writing will be limited from now on. Pupils have two bins to place their journals in each day, one labeled "I vant to be alone" and the other "Read all about it." For the former, Ms. Engle just peeks enough to verify that the pupil has made an entry, and checks it off. For the latter, she reads, answers questions or makes comments. After she checks off 10 entries in a journal, she gives them play money that they can exchange for some tangible reward like candy or teddy bear pencils. She says she does this to get them going on writing, "but some of them are now writing anyway," without the reinforcers.

Although they pause for the moment of silence, the kids keep writing through Dr. Thorne's P.A. announcements. He says that doing some research and telling him what made Molly Pritchard a revolutionary heroine will earn someone a compliment card. He praises one homeroom where everyone has made their day everyday so far this year. "For the first time, Manuel Garcia made his day. Manuel, keep up the good work, keep coming to school. Hopefully we won't have to go another 45 days for Manuel to make his day again." Announcing another in a continual stream of contests and competitions, Dr. Thorne says that there will be an Olympics of the Mind competition within Hamilton, and those who participate will also be getting practice in basic skills. The winners will go on to enter the district competition.

At 8:10 Ms. Engle tells the pupils to put away their journals and get ready for reading. In nearly complete silence, some of them get up and leave the room. While waiting for new arrivals, she rhetorically asks "What is it you are supposed to be doing?" Soon everyone pulls out books to read, most to do the book reports she requires. The form they use asks them to state the title, number of pages, number and names of main characters, synopsis of the story in their own words, and the reason they might recommend the book to a friend. Their parents are to sign the form.

During the transition, she has a quiet conversation with T, whose desk faces the wall and away from the class, almost like a permanent Step One in the ATF program. The other desks are arranged in three rows, the middle row consists of single desks facing forward, the outside rows are pairs of desks facing each other. She usually stands in the front of the room, next to her desk and in front of the board. She doesn't use the overhead as much as some, relying instead on the blackboard. The room is tidy without being military. Although there are no science exhibits in class, one of the boys has brought his pet python today, and it sleeps in a glass container. There is a bulletin board, labeled "Book It" that allows them to keep track of their book reports and later exchange them for pizza. Already in October, several kids have completed five reports; only one child has none. There are also children's drawings illustrating the books they have read: *Keeping Secrets, Ramona Forever, Snakes, Care Bears, Sammy the Seal, Space Case,* some of which Ms. Engle characterizes to a visitor as "baby books."

On another bulletin board are commercial pictures of clowns, labeled "We're not Clowning Around" and containing the chart of day-making; clearly fewer kids make their day in this class. Another bulletin board has Halloween safety rules. There is a poster of the six-day schedule, a poster calendar that lists assignments and tests (they are supposed to translate this into their own notebook calendars), and a clock on the wall that displays this message: "Time is Passing. Are You?" Like every classroom at Hamilton, there is a sign that reads, "You have no right to interfere with someone else's learning."

The transition complete, she now has 18 pupils, all in Level III, that is, below their grade level. She works with these children in two sets; one does seatwork while she directs the other in oral group work, then alternates the two groups. At 8:20 she calls one group up to the front. The students bring their chairs and form a tight circle around her. She addresses them in a soft but authoritative voice and uses humor and personal smiles. She tells them that almost everyone has earned his pizza certificate for October already and leads them in applause for one boy who has done five. "At the end of the year we'll have a big pizza party. That is your big reward for doing all those book reports. You have to do the book reports anyway, so you might as well get free pizza for it."

"We're doing Lesson 41 today. Make sure you have your books open to the right page. Sit up nice and straight. Books flat on your lap so I can see. Fingers on the words in column 1. The first word is 'expression.' What word?" Using a cricket snapper, she gives them the signal to recite the word. "Expression," they say word correctly in unison. "Spell expression," she says, and clicks the cricket 10 times to cue their unison response. "What word did you spell?" She clicks; they respond. "Good. The next word is..." She continues the exact routine through the words "remind," "couple," and "important." "Now let's read the words again." At repeated clicks the kids read the words consecutively, making no mistakes or self-corrections, their voices suggesting confidence. During the spelling of "repeat," one pupil reversed the a and the e. She detected it, and had the individual, then the group, repeat "repeat" several times.

As before, her manual never leaving the crook of her arm, she leads them through the vocabulary lesson. The program calls for the teacher to read the definition of a series of words they have already recited and spelled and ask them to recite back what that word means, or rather what the manual says the word or phrase means. For example, she says, "The next word is 'make a decision.' What does it mean to make a decision?" To answer correctly and receive credit the child must make a response consistent with the manual. "When you make a decision you tell yourself you will do something. "But when a child responds, for example, "make a choice," the teacher says that may be all right sometimes, but refers the child to the authority of the manual and the book in front of him. The responses must be as written in the book (even though others might be linguistically comparable or even superior) and must be in the form of complete sentences.

The next section is oral reading of words, then stories that contain those words. "If you read the words correctly you will earn two points in your WA box." As she calls on particular children (everyone eventually gets a turn), they read from columns of four or five words. If there are no errors, she says, "Good job. Two points." This proceeds rapidly. "Everybody earned two points in your WA box. Since you all did so well you earned one bonus point. You can record those points in your workbook now. Find Part B in your reading book. Backs straight, books flat in your lap, hands away from your face. What is the error limit for this section?" She clicks her cricket; they respond. "Six errors, right. There are extra bonus points for each error you don't make. So be sure you look carefully at the words and make sure you read exactly what's there." They only make two errors and reward themselves with bonus points. Then comes the story.

"The title of this section is 'Nancy looks for food.' What is going to happen, T?....Right, she's going to look for more food. What happened at the end of the last section?....Right, she was hungry." There are three questions put to the pupils about why we get hungry that are directly from text. "We're human." "To grow." Ms. Engle reads the next section aloud as the kids follow diligently in their books. Then she calls on children, one at a time, to read two or three each. She asks more questions to test their comprehension and memory. "What does it mean, she followed her nose?" "Why couldn't Nancy see very well?"

During the reading there are almost no mistakes, though children read one word at a time and use their fingers. When one misses on "with," Ms. Engle asks them repeat it several times and spell it, individually and as a group. Then the person who made the error repeats it several times. This was the pattern throughout the reading lesson.

Although this group has no trouble with the oral reading, in other classes children become tense as they approach the error limit for a passage. Even when they have been reading effortlessly and competently, they choke up when, by exceeding the limit, their error will cause their group to repeat the story. So group pressure is apparent. Not here, though, as this is a low group proceeding at a steady pace,
probably for the second time through this material. The degree of difficulty allows them to work with few errors, but neither does it extend them.

Again they get all points and bonuses, and Ms. Engle starts them working on their skillbook papers. "You all want to get 100 percent on your workbook. Is there anyone who doesn't want 100 percent? I want you to go back and recheck each answer. Make sure you wrote down exactly what was in your head. When you are done, work on your Halloween bookmarks." As before, the skill book worksheets call for them to fill in blanks with specific answers to closed-ended questions and cover the same material they covered orally in the group.

At 8:52, the students return to their desks and start to work on workbook questions that are exactly like those just asked orally. The work is easy, and everyone is done in 10 minutes and quietly working on book reports or bookmarks.

Later, the teacher has them grade their own papers, using a colored pen. She reads off the correct answers from the manual. When the response is variable at all, they raise hands to get a ruling. She verifies the correctness of most, but one case illustrates the authority of the program and the value placed on convergent responses. To the item, "Name three things that Nancy learned when she was small," one girl writes, "that her mother couldn't hear her," and asks whether she can count it correct. Ms. Engle replies, "That's true, but is that one of the four things we talked about in the story? So that wouldn't be correct." They finish and total their errors and put the score in the error box. Then she calls their names one at a time. They call out the number of errors they made, and she recites the percentage and marks it on the sheet. She praises them because almost all made two or fewer errors. To the one girl who made five errors, she says, "N, I will have to see you later. Bring your workbook." Without missing a beat, she assigns them Lesson 43, and they begin work. What they do not finish now they must be do as homework.

At 9:45, when members of Ms. Engle's homeroom reassemble, she takes points for ATF. She assigns everyone full points for homeroom. For the reading period, she recites the names and infractions of those who failed to earn full points. For three kids, she says, "You're down 20 points for not reading during reading time, not working on your journal, and playing with crayons during correction. J, you're down two points for not bringing a pencil. Everybody else, you're all set."

After going to music, which they do two days out of the six-day schedule, and a break when they could finish work or read books they choose, she conducts a spelling lesson, from *Spelling Mastery.*

"Get ready to spell some words that have more than one morphograph. Remember that a morphograph is every part of a word that has a meaning. That is the terminology we use in *Spelling Mastery.* Your other text uses the terms prefixes and suffixes. What is the first word?" She clicks; they respond, reciting in unison, "Misjudge." "Misjudge, right. What is the first morphograph in misjudge?" She clicks, they recite, "Miss." "What is the second morphograph in misjudge?" She clicks; they respond. "Spell misjudge." She clicks eight times to cue their choral spelling.

She follows the same routine exactly from the manual in her hand. On the words they miss, she has them repeat the word and its spelling individually, in groups, with and without her reciting with them, slow or fast. "Find Part A in your book. You're going to write the words that you just spelled. Oh goody. But this time you're going to get 100 percent, right?" She reads the words and uses them in sentences as they spell them in workbooks: uneven, restless, misjudge. "When you finish, put your pencils away and get your markers out." There is a noiseless search in their desks for markers. "The first word is replace. Spell it with me. Ready." She signals with the cricket; they respond. "Good. If you made any mistakes, write the word correctly. They repeat this routine for the rest of the words.

"Now put your marker away and take a pencil out and show me that you're really ready to go. This is your spelling test. There are 30 words that you have had before. You have spelled them orally and you have spelled them in sentences. Now I want you to spell them in the test. Make sure your brains are turned on and you write the word as it is in your mind. The first word is city. What word?" She clicks; they respond. "Write city." Other words, given in exactly the same pattern, are believe, equal, gold, truck, serve, want, real, and humor. The kids are implacable, showing no signs of nerves, except some erase their words to correct them with enough vigor to tear the paper. Some place their arms in a manner to suggest shielding their work from the eyes of others, but they could also just be holding down and steadying the paper. Occasionally someone asks Ms. Engle to pronounce a word. When they finish five minutes later they correct their mistakes and call out the number of errors they made. The average number of errors is two.

In Ms. Engle's class, spelling alternates with writing. A typical lesson from *Expressive Writing* is the following: The teacher presents a worksheet to the pupil with a picture on it, instructions, and a set of words to use. The model includes these words: talked, wrote, telephone, stool, celery, apron, numbers, piece, shoulder, paper, sat, clipboard, bunch, and bananas. Instructions are: "Write a paragraph that reports on what Robert did. Copy the sentence that tells the main thing Robert did (Robert worked at the store). Then make up at least two more sentences that tell what he did. Begin each sentence with he." The teacher reads the instructions from the manual and calls on a student to "Read the sentence that tells the main thing Robert did," and gives the signal. "Look at the picture and get ready to tell me another sentence that reports what Robert did. Start your sentence with he." The teacher calls on several students and praises responses that are correct according to the presentation manual. If a student says something like, "He is sitting on a stool," teachers must rephrase the sentence so it is in the past tense. Teachers lead the students through a unison recital of the words listed and tells the students that if they use these words, to make sure that they are spelled correctly. When students have written three sentences, the teacher has them check for errors of indenting, capitalizing, and punctuating, using present tense or not starting each sentence with he. The teacher calls on several students to read their paragraphs, providing praise and corrections as needed.

After lunch and recess, which involved a volleyball competition between the fourth-grade classes, she does a notebook check to see whether they have marked tests and assignments on their own calendars, but does no formal evaluation of this. Then J gets to take his python out of the case. The children take turns touching it. Although some ask what it eats or how big it gets, Ms. Engle does not seem inclined to make connections to natural science. Considering the potential for hysteria or involvement with this subject matter, the kids remain as quiet as ever.

After about 10 minutes of this it is time to take points again, and this takes 10 minutes. When visitors come to Hamilton to observe ATF, the principal usually sends them to Ms. Engle's room. They see the two parts of the discipline system. One part involves the accumulation of points toward, "Making My Day." For five different periods each day children assign themselves points they feel they earned for that period. They justify points by saying, "Did everything that was expected." To justify bonus points, they say, "Read a book," or "Worked on spelling," rather than taking a break. Then children may challenge the points others assigned themselves. To disagree, one child must face the person he challenges, state the reason, and suggest the number of points that ought to be assigned. The child who is challenged then must agree to accept the reduction or disagree with the challenge. The teacher adjudicates at this point or assigns the average of the two point values. At the end of the day, the separate points of the periods are added together to determine whether each child made his or her day.

The other side of ATF (Attitudinal Transitional Format) is "going on Steps" for breaking the sole rule at Hamilton, "No student has the right to interfere with another student's learning." Putting a child on Step One means that the teacher directs the student to sit in a chair facing away from the class. Then, after a period of time, the teacher asks the child whether he knows what he did wrong and whether he is ready to return to the class. When a pupil fails to respond appropriately at Step One the teacher puts the pupil on Step Two (standing facing away from the class). Step Three involves the pupil standing with his or her nose next to a dot on the wall. Step Four involves the child facing the wall in the principal's office until his parents come in for a conference.

An ATF Coordinator from the district office visits each classroom periodically to make sure teachers follow the procedures precisely. Nevertheless, classrooms differ. Some children collude with each other by rarely disagreeing with other children's assignment of points. In other classes, children seem to use ATF challenges as weapons in their social competitions. In still other classes teachers seem to collude with pupils, going through the procedures enough to satisfy the ATF coordinator when he comes through, yet maintaining relationships and classroom controls by some other personal or traditional methods. Ms. Engle, though, goes by the book.

This time through, the taking of points lasts 12 minutes. Pupils justify bonus points by saying, "Fifty. Violin." "Fifty, read a book." During disagreements, R challenges T because, "You were talking to D on the reading rug." T accepts a five-point reduction, then when it is his turn challenges R for doing homework during point-taking. Someone else challenges J's bonus points because he could not find a pencil. M disagrees with A because she offered no justification for her extra five points. M also disagrees with L: "L, you were talking on the reading rug." L has a logical defense so Ms. Engle gives L his 50 points but decreases M's points for poor choice of disagreements: "If you disagree, make sure you have a valid defense." Later, at the end of the day when they add up the points, five pupils will fail to make their day.

After another five minute break it is time for math. Since the fourth-grade teachers have elected to group pupils by math ability (they used ITBS and third grade CUES results), the children deploy to their proper room. There are some problems on the board that they start to work:

17 - 8 =

6 + 7 + 4 =

9 x 9 =

6/ 54 =

What does the digit 4 mean in the numeral 14,203?

This is part of what the district calls Systematic Review, wherein teachers begin math period with problems from an earlier unit in the math curriculum. They do this everyday to sharpen their skills. They work these silently and then she asks for volunteers to work them on the board. They make no errors.

Instead of their regular math lesson, which would have been solving computational problems in long division, she decides to give a CUES test. This requires them to back up over old material. Introducing the test, she says, "Yesterday we talked about some times, like 20 minutes before the hour, 20 minutes after." Several of the children groan audibly. "But that was pretty easy for you so we might as well do CUES now. When you get your CUES booklet, make sure you have your full name and your homeroom on it." Each pupil has a booklet with the CUES they will take during the current reporting period. Let's look at 4-3 first. On number 1 through 5 you're going to write the numerals. They have given you 5 ones, 6 hundreds and 3 tens. You are to write the number so the 5 is in the ones place, 6 is in the hundreds place and 3 is in the tens place. Okay? That's what you're going to do in 1, 2, and 3. In numbers 4 and 5, they gave you the number statement, and you are to write the number. Make sure you put the commas where they should be. If you look at the next section, number 6 through 12, what does the digit 3 mean in each numeral? You are going to write the place value. Number 13-16, show how to read the numbers. They give you the numbers, you need to fill it in among the words they gave you. Or, fill in the family name in the space. On 17-25, put the symbol for greater than or less than." In answer to a pupil's question, she repeats her previous words and tells them, "I can't give the answer for you, but if you look at the number closely you'll be able to answer your own question. Any other questions. Look at page 4-4. For 26-30 you are going to round those numbers off to the nearest ten. Circle the word ten in your directions as a reminder. Remember it's the nearest ten. Number 31-35 you are going to be rounding to the nearest hundreds or to the nearest dollar. Any questions on those? Now comes your most favorite part in the whole wide world. Numbers 36-39. All the information you need to answer those questions is right there in your chart. Remember to use the chart to figure out the answer. Number 40 right there at the bottom is another word problem." She continues giving instructions on different section of the CUES, and then they start to work.

Exhibit Two contains the CUES test the class takes.

Ms. Engle circulates, looking over shoulders. She reminds them, "Don't rush through it so fast that you're not thinking. Make sure that you are reading what they are asking." Another hand is up. She looks at the pupil's paper and nods.

There is utter silence, no outward signs of stress, a businesslike atmosphere, no wandering eyes. After looking over a shoulder, she says to the whole class, "If you get stuck on one, and you just can't figure out how to do it, don't keep at it for 20 years. Skip over it to the ones you can do and come back to it later."

On a section where the children must round off numbers to the nearest hundreds, one girl is rounding instead to the nearest tens. When Ms. Engle notices this she points out what the directions say. The girl erases and makes corrections.

After about 10 minutes, the first pupil finishes, but Ms. Engle stops her from turning in her test. "If you're done, recheck those that gave you trouble. When you're sure of your answers, bring them up to the table for scoring, then you can take a number search [worksheet]."

She looks over another shoulder and notes to her intern that almost all the children rewrite items presented horizontally and work them in the margins in vertical form. She says the horizontal problems "throw them. That is why we always include them in the review problems I put on the board. We try to get them used to that format before the ITBS rolls around."

About 25 minutes into the test, she announces that there are four minutes left before they have to finish their math period and return to homeroom. She notices that one pupil looks frustrated and in need of more help. Finally she goes and tries to help her figure out the story problem about distance between cities. "You're supposed to find the distance from New York to Paris. Where is it on the chart? Round it to the nearest hundreds. What is it?" He writes an answer and she nods. This problematic item for rounding is unnecessary to answer the question, and rounding makes the answer equivocal.

The intern who is correcting the CUES tells Ms. Engle that "pretty much they're all missing the same ones." She says "that tells me we need to cover that again." If too many pupils miss too many items, she may have to form a lower math group.

At 2:00 there is another silent transition, and the homeroom pupils come back. Again they take 10 minutes to take points, and the competition and revenge maneuvers from the earlier points recur.

Now it is time for science, and she calls for them to turn in their homework worksheets. There is an unpleasant scene as most of the kids turn in homework passes (which the ATF program allows as an occasional break from assignments) instead of their science homework. Peeved at what she sees as a frequent ploy for avoiding assignments, she says, "Here is a list of people who I won't accept homework passes from anymore. These are people who got notices that they were failing science. Those of you who do not have copy master 16 or a valid homework pass are receiving a zero in the grade book."

> She directs them to a page in their science books on classification of leaves. She quizzes them on how well they remember how the text classifies leaves, and then they take turns reading from the text. There are no leaves, flowers, or other flora in the classroom, nor are there pictures of them. Although they religiously follow along in the text as others read, the pupils display no signs of interest. Ms. Engle does science like everyone at Hamilton does Reading Mastery: unison reading or individual reading of two or three sentences in the text, choral reading and defining vocabulary words, answering questions from the text that measure memory for details, and comprehension of the material just presented. Although the HBJ text suggests some laboratory and field studies and activities, there seems to be little room for them in the weekly schedule. Although Hamilton has one of the finest laboratories of any elementary school in the observers' experience, few teachers other than the science specialists use it. Ms. Engle has different priorities, apparently responding to the principal's charge to emphasize language, reading, and math. Learning science and social studies depends on a foundation of basic skills.
>
> Several things about assessment come clear in Ms. Engle's class. All students must master a common content. Groups or individuals must recycle through this material again and again until mastery is perfect. Teachers may place those who lag behind into slower-moving groups, but pupils who do understand cannot accelerate. Recycling toward perfect mastery ignores the possibility that tests contain ambiguous items or that the motivation of pupils to read carefully and answer correctly may not be constant. Learning is a matter of correct performance, producing correct answers to closed-ended questions. In this respect, correct performance on worksheets and tests is analogous to correct behavior, in that teachers expect pupils to behave according to common rules and procedures. Thinking, to the extent that teachers consider it at all, consists of tricks of logic and analogy pupils may use to produce the correct performances. Assessment is simply adding up the number of correct performances and comparing them to a predetermined standard. Teaching, it logically follows, is correct adherence to the authority of the materials and texts. Finally, we learn that some forms of instruction are almost indistinguishable from assessment.

Testing in Stage Three. Teachers think less about external testing and more about ordinary instruction. Nevertheless, testing intrudes. Teachers use assessment to advance instruction. When they must assess in ways that they perceive as not advancing instruction or satisfying their educational goals, they perform the assessments in ritualistic ways. Tests play other roles in ways that teachers may not know about or acknowledge. The school requires teachers to cover so much curriculum ground that they must let some things slide. What they choose to let slide is apt to be those parts of the curriculum that do not appear on external tests. Teachers who neglect curriculum that is on the test or Scope and Sequence in favor of content or modes of instruction that they think are more educationally sound do so at some personal risk. Doing both requires enormous energy and more time than is available.

Tests also play hidden roles. When modes of instruction mimic modes of testing, teachers may not even recognize that they are teaching to the test merely by teaching. Instruction has already been aligned without anyone considering whether the change is beneficial. Similarly, when formative tests such as CUES mimic summative, external tests such as ITBS, merely taking the formative test is equivalent to practicing the summative one.

Stage Four: Planning for the Upcoming Test

Abruptly, after the winter break, the orientation toward testing changes from background to foreground in the thinking of teachers and is more evident in their actions. Ordinary instruction proceeds apace, yet directives and messages from the school administrators rivet the attention of the teachers to the tests. Increasingly, teachers pass along similar messages to students. What the students hear is, "This material will be covered on the ITBS, so be ready." Or, "They're going to drown you in tests!"

At Hamilton, a round of teacher evaluations occurs in late January. The primary function of these periodic "observations" of teachers and classes by the principal and assistant principal is to make sure that teachers are carrying out the reading, language, math, and discipline programs exactly as prescribed. The administrators schedule an observation, "script" the class, provide feedback to the teacher on strong and weak aspects of his or her teaching and suggest improvements. As later reported by several teachers, the secondary function of this round of observations is to inform teachers of the importance of attaining high scores on the upcoming ITBS. An intermediate grade teacher makes the following statement during a chance meeting:

> I think you should know that my whole way of thinking has changed since I last talked to you. When we talked, it was before my last observation, and at that time I told you that I do nothing different to prepare my students for the tests. But in the meantime I've had my meeting with Dr. Thorne, and he said that "The test scores are what the district is emphasizing. We may not agree with it, but this is the way we are going. So it's important for our kids to do well on these tests. So you need to shine." So from now on I'm going to have to think differently about it [test preparation]. First thing, I'll have to get together with [other teachers at my grade level] and find out exactly when these tests are given. Then I'll try to do everything I can to find out what's on the test. When I know what's covered I'll start doing some intensive reviewing. We already do Systematic Review in math, but now we'll do it on whatever the test is going to have on it. I'll scrounge around and find out if the district has any review material available and use that. Maybe it [what I do in class] won't be much different than it is now, but it will be interesting to see how it develops. But my approach has definitely changed.

A staff meeting at Hamilton in mid-January also illustrates the changing meaning and orientation to testing.

Exhibit Two

Name____________________________________ **Unit 2 Test** | page 1

Write the numerals. (Objective 2.1)

1. 5 ones
   6 hundreds
   3 tens

2. 8 hundreds
   1 thousand
   0 ones
   4 tens

3. 9 thousands
   0 tens
   2 ten-thousands
   0 ones
   7 hundreds

____________ ____________ ____________

4. 7 million 50 thousand 123 ____________

5. 40 million 563 thousand ____________

What does the digit 3 mean in each numeral?

6. 5,360 __________
7. 14,203 __________
8. 139,204 __________
9. 314,687 __________
10. 53,164,280 __________
11. 319,861,000 __________
12. 457,036,159 __________

Show how to read the numerals.

13. 5,043 = ____ thousand ____
14. 316,000 = ____ thousand
15. 70,300 = 70 ________ 300
16. 130,056 = 130 ________ 56

Obj. 2.1 ☐

Write >, <, or = for ○. (Objective 2.2)

17. 61 ○ 59
18. 87 ○ 101
19. 335 ○ 342
20. 526 ○ 527
21. 1,201 ○ 978
22. 41,325 ○ 41,325
23. 649,261 ○ 649,620
24. 26,403,178 ○ 26,403,177
25. 310,010,000 ○ 310,101,000

Obj. 2.2 ☐

Macmillan Publishing Co., Inc.

Exhibit Two

Name______________________________ Unit 2 Test | page 2

Round to the nearest ten.

26. 63 27. 405 28. 184 29. 96 30. 294

Round to the nearest hundred or nearest dollar.

31. 438 32. $6.52 33. 861 34. $1.48 35. $3.60

Solve the problems. Use the distance table for Exercises 36-39.

	Flying Distance (in kilometers) Between Capital Cities			
	London	Paris	Rome	Washington
London, England		346	1,435	5,915
Paris, France	346		1,107	6,180
Rome, Italy	1,435	1,107		7,234
Washington, D.C., USA	5,915	6,180	7,234	

36. How far apart are Paris and Rome?

37. Which two cities in the table are farthest apart?

38. The Smith family flew from Washington to London. The Horn family flew from Washington to Paris. Which family flew farther?

39. Round the distance between London and Paris to the nearest hundred.

3

40. Ms. Holt keeps a record of the cars that are rented OUT and those LEFT on the lots of a company. Complete her record.

CARS	AT START	OUT	LEFT
LOT A	10	6	
LOT B	14	9	
LOT C	9	3	
LOT D	8	8	

January Staff Meeting at Hamilton

> The meeting follows a typical pattern, almost all teachers gathering in the library after school, tired but still with a sense of humor. There is a written agenda, and Dr. Thorne shares the task of presiding with Dr. Michael. They announce upcoming events—a special program on birds of prey, an actors class for the children, and a staff development day where teachers will discuss long-term goals for the school. Then Dr. Michael speaks about the first of the external tests on the schedule, the district Study Skills Test. "The test will be multiple choice. It will be more of an application type of test. I see this as a very important assessment. And I see us needing to do very well in grades three through six. I really want to see some emphasis on what you do so that when you take the test the kids are prepared. The first year we administered the test, the school averaged 86 or 87 percent, and I'd like to see us have the same success rate or better. You'll need to do some heavy review for this test."
>
> Dr. Thorne then anticipates the ITBS. "Turning to some other assessments, it would be remiss of me as principal not to talk about student achievement and some of the testing going on. This has been an unusual year for us because we were staffed for 825 students and we're currently at 717. So we are operating, supposedly, with more staff than we should have had. And, consequently, this year our course loads have been a little bit less. It's very important, politically, that our kids come out well on the ITBS—as well as they can. And that they perform well on the Metropolitan. And given the kind of kids we deal with, they will need to be prepped pretty well in taking the test, especially on reading and math. I don't need to say a whole lot more than that, and Basic Skills Test the same thing. It's important as I sit down with the superintendent and those powers that be that I've got something in my hands that I can deal with. Our kids are achieving in terms of the rate of growth. That really works for me in terms of dealing and getting the things I've been able to get for the school. So you know where I'm coming from on this. You've worked very hard. You've continued to emphasize academics in instruction and that will show up on the test. The kids have got to do well. You've really got to pour it on."
>
> The ESL teacher asks whether the ESL kids will have to take the test, and Dr. Thorne says they do if they've been here in school more than a year, otherwise not. She says with finality, "They will flunk them all." He shares her concern and regrets that the state lumps all the kids together for the purposes of testing. But, in spite of the drawbacks, "we CAN make growth," and demonstrate growth from year-to-year, the district goal is to produce one grade equivalent month gain for every month spent in school. "Enough said. You know it's important. We've GOT to pull through."

The message the principal gives to teachers is a transmission from the district office. Recently, the schools received a memo reminding them of the importance of attaining ITBS growth in excess of the growth attained last year. It asks the principals to estimate the amount of growth they expect the school to accomplish. Cactus administrators are well aware that other districts that failed to pass bond elections were those with low test scores. No doubt they look at this correlation and wonder whether a few ITBS growth points might result in loss of hundreds of thousands of dollars in the next election.

About this same time at Hamilton, certain teachers ask for permission to bail out of *Expressive Writing* and *Writing and Thinking* because of the discordance between the content of these programs and the format and the content of CUES, BST, and ITBS in language arts. Third grade teachers issue a formal complaint that, with all the programs that Hamilton requires of them in addition to what the tests cover, it is just too much. The teacher who elected to use Math Their Way for the lowest-performing third grade math pupils has already given it up in favor of the regular math curriculum of memorizing math facts, subtraction with regrouping, and speed drills on simple addition and subtraction problems.

The change in orientation toward testing is also evident at Jackson during grade level meetings in mid-January. It is here that the intermediate grade teachers are becoming aware of the disparities between ordinary instruction and pilot CUES, and the format and content of the Basic Skills Test for which the district holds them accountable at the end of the year. The primary teachers are concerned about the disparity between Math Their Way and the skills covered by CUES and ITBS. How will they bridge the gap between the concrete and the symbolic? Among themselves, teachers discuss ways of coalescing ordinary instruction and coverage of skills the tests cover. Also within grade levels, teachers begin to discuss what they will do to prepare their pupils for the ITBS. They mention several possibilities. Some reject making any alterations in teaching to accommodate the test. "At the school where I taught last year," says one primary grade teacher, "we started doing worksheets to prep the kids for the Iowa way back in October. Now I realize this is a cop-out. If I really feel that the test is not valid and the scores aren't important, I have to just keep doing what I'm doing and not worry about how well they do."

An alternative plan is to use the materials, *Math Objectives Review* (MOR) and *ITBS Correlations*, which the district publishes to help teachers prepare pupils for the ITBS. The intent of the materials is to provide some familiarity with item formats and to review skills in the district curriculum. The third and most controversial option is to use *Scoring High on the ITBS*, a package of worksheets and materials that mimic ITBS items. Certain administrators from the district and the state department discourage use of this package. Jackson's teachers suspect that if they use *Scoring High*, people will think that *Scoring High* was responsible rather than Whole Language. Mrs. Mitchell is ambivalent, and she gives no directions one way or the other about test preparation. Nevertheless, a limited number of copies of *Scoring High* are available, and many of the teachers study them carefully and adapt them so as not to violate copyright laws. Those who favor the use of *Scoring High* argue as follows: This package, because of its close resemblance to actual test questions, has the chance to raise scores. By using the materials for an hour or so a day, I can lessen the impact on ordinary instruction while increasing my average test scores.

Testing at Stage Four. This stage in the natural history of the testing event is one in which teachers begin to orient ordinary instruction to the contents and formats of upcoming tests and make plans to prepare pupils for the tests. Principals give mixed messages. They want scores to be high, but they hesitate in directing teachers to prepare in particular ways. The messages principals receive from the district mention raising scores but not enhancing educational goals as the means for raising scores.

Stage Five: Putting Ordinary Instruction in the Background, Test Preparation in the Foreground

Test preparation takes several forms, depending on the external test. For example, the first external test of the year is the Study Skills Test (SST). To prepare pupils to take this test, teachers borrow time from other subjects, usually social studies, science, or writing, and perhaps math. When the SST is past, so is instruction in study skills. Teachers revert to ordinary instruction until it is time to prepare for the next external test. During this stage, ordinary instruction fills in the spaces around test preparation.

Study Skills Test. For the three weeks between the staff meeting at Hamilton when the principal urged the teachers to prepare, and the week of SST testing (the week of February 8), teachers alter their routines.

The district has a required curriculum in study skills for all grade levels, but does not require testing until fourth grade. Testing third graders is optional for each school in the district. In primary grades, the skills consist primarily of organizing materials, following classroom rules and procedures, and writing names in the correct place on the paper. In second and third grade, teachers follow a manual to instruct pupils how to use reference materials. In intermediate grades, pupils have a handbook that district staff prepared that covers 30 to 40 short topics, including checklists for organizing materials, maintaining a calendar of assignments, tips on completing assignments and chapter questions, proofing assignments, studying for different kinds of tests, tips for taking different kinds of tests, active reading, note-taking, tips for turning the question on a test or worksheet into part of the answer, the parts of a book (index, etc.), graphs and charts, use of the dictionary, and the technique of "RCRC" (Read, Cover, Recite, Check). Many of these topics and skills are also covered in the prior grades. For example, the skill of using "guide word" to find words in the dictionary is covered in every grade from second to sixth. More than once, Dr. Thorne has referred to the study skills curriculum as "survival skills, especially for these kids that we have here." According to district guidelines, teachers must make regular progress through these materials during the school year that is a part of ordinary instruction. Then comes the test, which is a mastery-level test of the competencies. Preparation for the test consists of repetition of the exercises the handbook covers. The following vignette shows a typical pattern of preparation for the SST at Hamilton School.

Mr. Armstrong Prepares for the SST

> Coming into Mr. Armstrong's classroom is almost like entering a conservative men's club. Lighting is indirect. Classical music plays softly in the background. There is a reading room off to the side, set off by L-shaped bookshelves, where students can read on comfortable sofas or gather around a circular table for conferences. Like Mr. Armstrong himself, the room is orderly, nothing out of place. By this time, Mr. Armstrong's sixth graders have already completed the series of lessons in the district study skills handbook. Mr. Armstrong agrees with Dr. Thorne about the importance of

these skills, and repetitive review fits his conception of how pupils learn and remember. He endorses the SST because, unlike the ITBS, "it was written by the district, by the same people that wrote the workbook."

Although the class has completed the workbook during the fall term, Mr. Armstrong has taken seriously the urging of Dr. Thorne to make sure the pupils in his class do well. He uses the initial 15 minutes of the day for silent review of the workbook. Previously, before the push for test preparation gets started, his class occupied this time writing, reading library books, completing assignments, and listening to the morning's announcements over the public address system. At 8:00 a.m., following this relatively unstructured time, the class disperses to "specials," either physical education, music, or art. When the class returns at 9:00, the students normally spend the next 50 minutes in math. For the three weeks before the test, however, Mr. Armstrong uses not only the first 15 minutes of the day but also diverts anywhere from 5 to 15 minutes of math period to review for the SST. One might think that he could just move all the subjects to later in the day, but at Hamilton, there is almost no flexibility. No one messes with specials, and the period devoted to reading instruction is sacred. At 9:50 a bell sounds, and pupils all over the intermediate grades disperse to their reading groups from 10 to 11:30 each day. In Mr. Armstrong's class, four pupils leave and six others come in from other sixth grade classes, so reviewing for something other than reading must fit around this schedule. Occasionally, Mr. Armstrong takes time away from the language class period rather than math to review for the SST, so the damage to ordinary instruction in math will be less.

At 9:05, about a week before the test, he announces, "Let's take five minutes for the study skills. That test is coming up soon. So before that...we've already gone through this book once, so now we'll start reviewing it. And continue reviewing it for a few minutes each day. And when we get to the end, we'll go back over and review it again for five minutes each day. So someone tell me when it is 10 minutes after 9. The next time through, we'll work some exercises that we didn't do before."

"We left off on page 52, on bar graphs [see Exhibit Three]. The only difference between this bar graph and the other bar graphs we covered are that these are horizontal and the other ones are vertical. I'm not sure it is so important to discuss the difference between the direction the bars are going, but let's take a look at what they represent. What is the title of the first graph on page 52, M?"

M doesn't answer so he calls on C who has her hand up and reads the title, "Area of the World's Largest Nations."

Armstrong: "So that is what the graph is trying to picture. Now, going down the left side of the graph, J, what is trying to be presented by those bars going across?"

J answers correctly the question that Armstrong stated, but not the question Armstrong intended, which is "what do the words represent on the left side of the graph?" J says, "Millions of square kilometers."

Armstrong: "That's not so, down the left side, what is being represented?"
J: "Oh, the names of the nations."

Armstrong: "Yeah. The names of the nations: the Soviet Union, Canada, China, the United States, Brazil. Okay. Then, the business across the top is the millions of square miles. Notice that they have something different across the bottom, that is the millions of square kilometers. So that's kind of interesting, how they can show that kind of thing on a bar graph. So if I ask the question, which of the nations is the largest in area, you could know right away just by looking at it. What's it going to be, E?" E answers "Millions of square miles."

Armstrong: "Listen to what I'm saying. Which is the largest nation in area?" This time E responds with the correct answer, the Soviet Union. Armstrong: "Yeah, you would know just by looking at how far the bar goes across. That's the plus for bar graphs. S, how many millions of square miles does Canada have?" S says, "About four?"

Armstrong: "Yeah, about four. If you wanted to get fussy, you could cut that in half, and that would be about three...million, and maybe five thousand. I guess the easiest way would be to say it just like you did. The chart underneath is the five nations with the largest populations. And which country has the least in population?" He calls on Ch, who raised his hand, and Ch answers "Indonesia," with a questioning inflection at the end of the word.

Armstrong: "Yeah, Indonesia. The number across the top represent millions of people. Next we come to line graphs on the next page. This is something you could do yourself with your grades. That would be kind of interesting. The first one here is talking about temperatures in Copenhagen, Denmark, for some reason. And down the left side is listed the degrees of temperature in Fahrenheit. Down the bottom are the months of the year. The solid line represents the highs and the bottom line represents the lows. So. When they were making this chart, they would simply make a dot, they averaged out the temperatures and made a dot for the high and the low, and when they got done they drew a line to connect the dots. And that makes them see quickly and practically how the temperatures go." He demonstrates this by drawing a similar line graph on the board, draws dots to correspond to average degrees and connects the dots. The kids are looking at them, but it is difficult to tell whether they are involved. He shows how they could do the same thing with temperatures in Phoenix and asks them what month the temperature would be at the highest average. N says July. He says, "Probably more like August."

He goes on to show them how they could construct a line graph of their grades. He draws a prototype on the board, saying the scale of their average daily work would go from "God forbid, zero; to 100 percent." He writes abbreviations for the months across the bottom of the graph. He constructs some dots representing averaged percentages for each month, then connects them with a line.

At 9:12, he notices that he has gone seven minutes longer than intended and says with despair, "It's gone too long! It's gone too long!" He asks them how to figure out the average and gets several volunteered incorrect answers. Ju tells him the correct way, and he demonstrates with the daily scores, in percentages, of 80, 90, 76, and 75—averaging 80. "When you get done, you can see graphically how you did. If you had enough colors you could put your language grades in red on the same graph. That's the value of the line graph. But probably it would be confusing to put more than two or three subjects on the same graph.

At 9:15, he says, "That's that. That's enough for today."

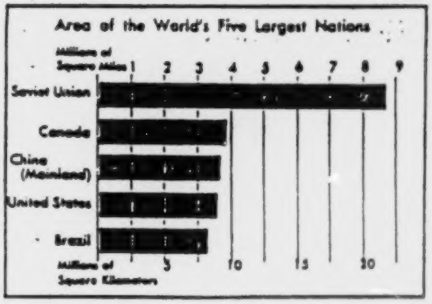
The next day, Mr. Armstrong also spends three times longer on review than he intends. When the pupils come in from specials, they seem to know that study skills is first, as most have their notebooks and study skills handbooks open. "Let's see what new and exciting things are in store for us in study skills....At the staff meeting today they're going to give out the tests and then we'll take it tomorrow. They're the bubble-in kind of answer sheets. I hope we don't have to put a lot of garbage on it [surplus identifying information, etc.]. I'm not worried about THIS one because the district has put together the test and they put together the booklet so the test should be pretty straightforward. I'm expecting some pretty high grades. Let's continue reviewing. Page 8. I don't know how to make this clear other than just reading it. 'Studying for content area tests. Step One. Begin studying early.' That's good advice; we should all take it, but it's more, 'don't do as we do, do as we say.' Get the materials that you need to study. 'Step Three. Study. Use RCRC.' Boy, we'll never forget that! Some of you who made below 80 percent on the spelling test should try to use this. If you use RCRC, you WILL learn." He continues reading and interspersing comments. "'Studying for skills based tests. Step One. Begin studying early.' Same thing....Write those problems out. Come by the right answer yourself. This is a good hint for all the tests that are going to come pouring down on your heads in the next few months." He continues reading from the sections on taking tests. "'If you have difficulty on an item on a test, go on to the next one.' The danger in laboring over one problem is that there may be problems at the end that you can do, but you won't have time because you spent it all on one you can't do. When you're taking a test, answer the ones that are obvious." He reminds two girls who are leaving for band to review tonight for tomorrow's test. "On multiple-choice tests, usually take the first impression. I'll ask myself a thousand questions [second-guessing an item] and end up getting it wrong when I would have been better off sticking to my first answer. Our brains work wonderfully on an unconscious level....'When you have completed the test, go back and double check.' Triple check if you have time."

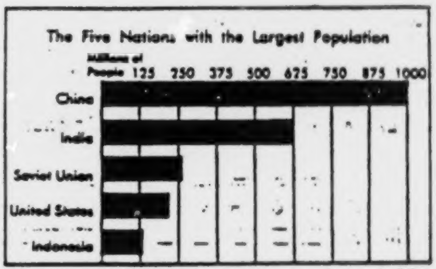
He continues reading, emphasizing the use of a process of elimination of obviously wrong choices on multiple-choice tests. "Make a light pencil mark next to the ones you can eliminate....On true-false tests, use that wonderful computer that we have. On essay tests—what is an essay test?" Someone answers, "It's where you tell what you know." He reminds them not to spend too much time on any one item, write down key ideas, and say the question in your own words.

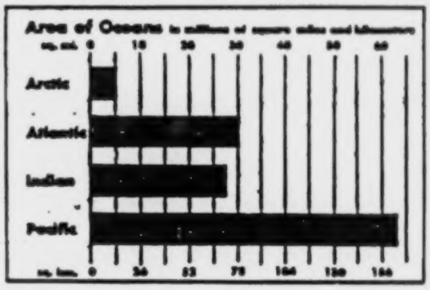
Exhibit Three

HORIZONTAL BAR GRAPHS

A bar graph shows us how things compare by the length of the bars. These are horizontal bar graphs.







Source: *Exploring Our World: The Americas.* Follett Publishing Company, 1980

© Skills for School Success, 1985

Exhibit Three

INTERPRETING HORIZONTAL BAR GRAPHS

DIRECTIONS: Use the horizontal bar graphs to answer these questions.

Find the bar graph labeled *Area of the World's Five Largest Nations.* This bar graph tells the area of the five largest nations in millions of square miles and kilometers.

1. Which country has an area of almost nine million square miles? ______
2. Which country has an area of almost four million square miles? ______
3. Which country has the most area? ______
4. Which country has more area, Canada or the United States? ______
5. Which country has more area, the United States or Brazil? ______

Find the bar graph labeled *The Five Nations with the Largest Population.* This bar graph tells the population of five nations in millions of people.

6. Approximately how many millions of people live in China? ______
7. Approximately how many millions of people live in India? ______
8. Approximately how many millions of people live in the Soviet Union? ______
9. Which of these nations has the largest population? ______
10. Do more people live in the Soviet Union or in the United States? ______
11. Do more people live in China or India? ______

Find the bar graph labeled *Area of Oceans.* This bar graph tells the area of oceans in millions of square miles and kilometers.

12. What is the approximate area of the Arctic Ocean in millions of square miles? ______
13. What is the approximate area of the Atlantic Ocean in millions of square miles? ______
14. What ocean is the largest? ______
15. Which is larger, the Atlantic Ocean or the Indian Ocean? ______

© Skills for School Success, 1985

> This is the third repetition of this material. E is yawning. M and A are staring off into space. N, as always, tries to participate. Reading their reactions, he says, "I don't want to spend any more time on this. Although it is dull and boring, these things really work."
>
> At 9:16, he tells them to put away study skills and take out math.

Although Mr. Armstrong's immediate goal is to enhance the SST scores of his class, his review also will likely affect their ITBS scores. By merely examining the contents of the handbook, one can see the correspondence to the ITBS, particularly in the repetition of graph skills and the instruction in techniques of taking tests. The content and format of the Study Skills Handbook mimic the ITBS, which is also evident in the following description of how Mrs. Wright prepares her third graders for the SST.

Mrs. Wright Prepares for the SST

> In her third grade class at Hamilton, Mrs. Wright also devotes about 15 minutes per day preparing for the test. The principal activity involved in this preparation consists of practicing dictionary skills.
>
> At 10:30, after the 90-minute reading period, 30 minutes of Spelling Mastery and a brief period in which Mrs. Wright reads to the children, they begin their lesson. Each pupil takes out a dictionary, and she asks a series of questions to individual pupils. "Where did you have to go to find the M's?" "Which way will you have to go to get to a P page? Front or back?" She has them look at the guide words in the dictionaries. "Look at the top of your page....There's some heavy black writing at the top of the page. There are two words." She then asks some students what the key words were in their dictionaries. She explains that key words are the words in heavy black print. She writes some of the key words on the board and says, "These are called guide words. They tell something about this page." Mrs. Wright then explains which words would be found between "palpitate" and "panel" (two of the guide words she had written on the board). She also reviews the direction in which the children would have to turn pages to get to a word that began with a different letter: "If you're on the word 'gum' and you want to go to 'victory,' which way will you go?"
>
> She then tells the pupils to open their study skills handbooks, and together they go through 10 exercises in determining whether (if they are currently looking at a word like "crazy" and want to look for a word like "gruff") they should go forward or backward in the dictionary. Already knowing the difficulty of the words, she comments on several occasions, "It doesn't matter if you can't read the words. You don't need to know the words to do it, you only need to know the alphabet." She then has them work several other examples independently.
>
> The next day the lesson is much the same, except that the words from the dictionary have the same initial letter. The exercises involve determining the direction to proceed going from word pairs like "pickle" and "panther."

> Later that day, the teachers get a message from the principals saying that they have decided to exempt third graders from the Study Skills Test. Expressing relief, the teachers abruptly drop study skills from the daily schedule. Although they feel that skills in using reference materials are important, they have too many other things to do.

At Jackson, most teachers place little emphasis on preparing for this test. Generally they agree that such skills are important, but that they ought to be covered "in context." Pupils ought to be able to extract such information from dictionaries and other reference books that satisfy their curiosity or are useful to them in writing or investigating particular topics. Sending pupils to the dictionaries, encyclopedia, and librarians are routine parts of unit study at Jackson. Intermediate grade teachers actively incorporate study skills lessons using graphs, tables, and charts into math and social studies. Primary grade teachers use graphs and charts as part of Math Their Way. Preparing for the SST departs very little from ordinary instruction in Jackson. Some intermediate teachers merely hand out the workbook and ask their pupils to work through it independently. Several teachers mention that, unlike the ITBS, there is no pressure associated with scores on the SST, and therefore preparing for it occupies little of their time and attention.

There are a few exceptions. Mrs. Grant, a fifth grade teacher, reports doing some systematic preparation for the SST. She makes up facsimiles of tests that resemble the SST items and administers them in worksheet and practice test format. But every time she asks them to open their study skills workbooks, they complain loudly about the boring and repetitive nature of the assignments. "We have to do these every year," whines one girl. Mrs. Grant worries that their negative attitude toward the workbook will carry over to the SST itself and result in low scores for her class. She tries everything she can think of to motivate them, relating the material to ordinary lessons and jobs, and showing the importance of the skills. But looking at the Study Skills handbook at several grade levels, one can see that the children are right. The district has overlapped skills and content. In pursuit of mastery, the district has achieved repetition and boredom.

Reflecting on the SST, a teacher at Hamilton provides this history: The state mandated the CUES so that the district would teach material that the ITBS covers. The state then instituted the BST to ensure that pupils remembered the material on the CUES. The district now requires the SST, so pupils are sufficiently test-wise to pass the BST and ITBS. From this teacher's point of view, the tail wags the dog.

BST-placement. Early in March, the sixth grade pupils take the Basic Skills Test in reading and math, which junior high teachers and counselors use to place them in a hierarchy of tracks. Preparing for this test consists of the teachers ensuring that they have covered the Scope and Sequence and the contents of the tests, which they have known since the beginning of the year. In Mr. Armstrong's class, for example, preparing for the BST follows similar patterns to preparing for the SST.

ITBS. Among all the external tests, the ITBS is the one that most displaces ordinary instruction. For both schools, but varying among teachers, the test preparation activities begin about two months prior to the April test week, intensify in more or less linear fashion, and continue even through the week of testing.

Teachers prepare for a subsequent test even on the day that another test is being taken.

From notes taken in observation of 6 classrooms as well as statements taken from interviews with 19 teachers, we recognized four types of preparation for the ITBS. In this section of the report, we describe and interpret classroom activities from each of the four types.

1. Reviewing Content of Ordinary Instruction

In this type of preparation for the ITBS, the teacher assumes a connection between the content of ordinary instruction and the content of the ITBS. Having made this assumption, the teacher retraces the curricular ground already covered. This amounts to intensive and repetitive review of material in texts, reading programs, and the study skills workbook. The latter, though a part of ordinary instruction, mimics content of the ITBS and BST in topics such as interpreting graphs and maps.

Within this type of ITBS preparation, the primary deviation from the established curriculum is a shift in the sequence of topics the class ordinarily covers. When the sixth grade teacher becomes aware that the ITBS covers concepts and skills of geometry, he moves up that unit from its place in the textbook sequence. This move supplants pre-algebra skills, which the test covers less extensively, until after the test. But in the period of rest and recuperation from testing, few topics are covered with the same vigor as before. Concentrating on the tested content, teachers who practice this variety of test preparation pay relatively little attention to coaching test formats.

Mr Armstrong Prepares for the ITBS

> It is one week before the start of the ITBS, but the week's activities vary only by degree from the previous four weeks. The pupils assemble for class (unlike pupils in the lower grades, whom the teachers must collect from the playground) before 8:00 a.m., open books or complete homework, listen to the announcements and any orienting comments from Mr. Armstrong, and then leave for special classes. There are now 17 pupils, though the size and composition of the class have been in almost constant flux. Only 10 of the children here now were here at the beginning of the year. While they are at special classes, Mr. Armstrong usually corrects papers, records grades, plans his day, and enjoys his coffee. Back at 9:00, the pupils reopen their "free reading" books, or work on the Systematic Review problems that are on the board. The district has recommended the Systematic Review as a way of reinforcing skills learned earlier, and the problems are ordinarily quite easy. Mr. Armstrong collects math homework, comments on the paperback book a boy is reading: "*The Entry*? How wonderfully demonic." Many of them have a penchant for Stephen King and other thrillers, but, he thinks, at least they are reading. At this point, three children leave for ESL classes, to reappear only before lunch and at the end of the day.
>
> On the overhead is a set of problems of changing fractions to decimals, a topic they covered a few months earlier. The first fraction is 3/5, and a boy calls out the correct answer. Mr. Armstrong complains that, "You're giving me

the answer. I want the process. Don't get ahead of me. Put your sign in this circle. Set up your division. Now work your problem. See how unbelievably simple that is?"

R, a 14-year-old ethnic minority whom Mr. Armstrong had earlier brought up to TAP, raises his hand and says, "I don't get it." Mr. Armstrong replies, in a matter-of-fact tone, "We'll do a few more, so you'll have a chance to get it. After that, we won't do anymore. So there is a point at which you have to get it."

He assigns them to work on their own a few more problems of this nature, taken from the Macmillan textbook. He calls them individually to his desk so that they can show him their answers. If someone works a problem incorrectly, he shows them the correct method and watches while they work some more, until he satisfies himself that they can do the problems successfully. These sessions come complete with lavish and genuine praise for good work and equally genuine reproaches for what he considers poor effort or lack of concentration. To R, who has not finished his homework, he publicly notes, "You're not coming across with the goodies. Do we need to be thinking of another move for you?" They all know he is referring to the sixth grade transition class. Notwithstanding the gruff and public character of these rebukes (which he levels against all of them at some time during the year), there is whole-hearted affection between Mr. Armstrong and his students, displayed repeatedly and incontrovertibly.

About halfway through the math period, several pupils leave for band and one returns from orchestra. Mr. Armstrong's indignation is apparent. "Another interruption. Someone should do a study about how many interruptions teachers have to endure." Even in this busy season, when the teachers feel pressure to get the pupils ready to take the tests, there are many extra programs scheduled: one on birds of prey, another celebration for Arbor Day, and a program on the solar system.

At 9:50, there is a transition to reading period; three pupils leave and four come in. They are working on spelling today, as they are near to completing Reading Mastery VI, and Mr. Armstrong feels more confident about their competence in reading than math, spelling, language arts, or the other content covered by the ITBS and BST. They spend the time working independently on a spelling lesson in the book, and he holds individual desk conferences as he held in math. He chuckles at the writing (each spelling lesson has a small writing assignment), and calls another "garbage." He suggests that an employer would not accept anything like what the girl had written there.

From 11:15, when one pupil asks to leave for a student council meeting (another interruption!), until 11:30, lunchtime, they take turns reading aloud from *The Return of the Indian.*

At 12:20 they reassemble, only to change rooms for science, coming back an hour later for language. For the first half of the year, Mr. Armstrong had used this period for writing. Now, however, he has turned their attention to the language arts book (Harcourt Brace Jovanovich) to review and re-review the points of grammar. Today's lesson is on singular and plural forms, and the object is to memorize the rules. He asks a girl to go to the board and "review those plural rules for us." She goes up, and from memory, writes:

-s	ch, sh, x, s, z	cy/vy	co/vo
dog	buzz/buzzes	city/cities	heroes
dogs	bush/bushes	boy/boys	radios
	fox/foxes		

f/fe	w/in	same word
shelves	tooth/	deer
wives	teeth	

Mr. Armstrong goes over examples for each rule and asks them to copy the chart onto their own papers. He then proceeds with a similar review of possessives, singular and plural form. After about 50 minutes, he assigns them a worksheet from the social studies text, and this, plus the usual ATF and housekeeping duties, accounts for the rest of the school day.

Test preparation activities are not as feverish this day as they often are in Mr. Armstrong's class, but they do illustrate a variety of test preparation. Teachers repeat material and skills they have already covered in ordinary instruction as a way of raising scores on tests congruent with ordinary instruction.

The principal effects of this type of test preparation are the following: Students become more familiar with the material that they review repeatedly. To the extent that the test covers this material, scores should increase (other things, like levels of effort, being equal). Real achievement on what is intensively reviewed, therefore, likely increases. The damage is done to other subjects. For example, in Mr. Armstrong's class, instruction in social studies grinds to a near halt. Science persists, but not to the same extent and quality as before; for example, the pupils go less often to science lab. Instruction in language changes character. Whereas in the fall term the pupils have many opportunities to write and have their writing critiqued, this segment of the ordinary curriculum has vanished, not to reappear until after the BST is over in early May. In place of writing, language arts consists of formal grammar and spelling from the textbook, emphasizing the rote memorization of rules rather than the use of language to express ideas. Finally, test preparation in this class has negative effects on progress even in those subjects that the test purports to measure. Time the teachers spend in preparing for the test might have been spent teaching lessons in math, reading, spelling, and language beyond that which the teachers are able to cover in the time they have.

2. Boosting Confidence

Ms. Anderson, a sixth grade teacher at Jackson, represents the type of test preparation aimed at boosting the confidence of the pupils to take the test. Although she uses worksheets from the district test preparation materials and *Scoring High*, her intent in using them is to prove to her pupils that they are smart, know a great number of things, and are capable of doing well on the test. Like some other teachers, Ms. Anderson uses these worksheets as vehicles for working on her pupils' feelings of self-efficacy. The test preparation activities in her class accentuate neither test formats nor the contents of tests or ordinary instruction.

She differs from her sixth grade counterpart at Hamilton, Mr. Armstrong, who assumes that the test and ordinary instruction in reading, math, and language arts are congruent. Her own reading instruction consists of literature study, and her writing instruction follows the district handbook, *Writing and Thinking*, neither of which coordinates with the district Scope and Sequence or the contents of the ITBS. In math, she is part of a pilot program using a mastery cooperative learning program called Team-Assisted Instruction. Although she supplements her own programs with materials the district developed in language arts, metrics, and reading maps and graphs, it seems obvious that intensifying ordinary instruction would not help her pupils on the ITBS. As much as any teacher we studied, Ms. Anderson follows her own lights and resists substituting definitions of curriculum laid down by the district or implied by test-makers.

Nor is she much interested in boosting the performance of her pupils by coaching them on test formats or teaching them content and skills likely to show up on the ITBS. She wants them to do well on the tests, not so much to make her look good as a teacher or to make the school look good, but for the pupils' own benefit, to make them feel successful. Although she spends large amounts of class time on the test-preparation material, the use of time is nowhere near as fervent as test-preparation activities in other classes we observed. Sometimes she assigns the worksheets as homework, sometimes as class work, afterwards telling the pupils the correct answers and asking how many they missed. Although she reads them the tips that are meant to induce test-wiseness, the reading is casual, almost as an afterthought.

Besides following through on her math program and pursuing literature studies and studies on units such as mythology and Egypt, her concerns seem to center on the adjustment and psychological well-being of her pupils. On many occasions she expresses her sympathies for them, particularly for how hard they have to work and the number of tests they must take. She repeatedly promises them incentives—parties, outings, treats—as compensation for their efforts on the upcoming tests.

Ms. Anderson Prepares for the ITBS

> Even with the ITBS only seven days away, the atmosphere in the class is informal, relaxed, unstructured. The pupils are taking advantage of a free period to read in books of their choice, write, and visit with each other. Ms. Anderson is holding a series of private conversations with each one about his or her choice of program for seventh grade. Her tone and body language suggest sincere concern for each individual, and her questions and comments are almost all about their feelings about seventh grade and the choices they are making. Ma, for example, had scored high enough on the BST to qualify for advanced science. She has left the choice up to him, justifying his eventual choice of the middle track because he felt more comfortable there.
>
> After an hour spent in attending to their concerns, she directs them to a science lesson, which involves reading a section from the text on color and light and completing a worksheet from the textbook program. She gives

them a choice of reading silently or aloud, and they vote for silent reading. During this month of test preparation, Ms. Anderson has attempted to rotate among subjects those that they will omit each day so as to slight no single subject entirely. The previous day, they had substituted test preparation for science, and today it will be social studies that is skipped. They work on science for the next 20 minutes (whatever part of the worksheet that they fail to complete they must do at home), and she dismisses them for physical education.

When they return, she spends 15 minutes discussing with the class its loss of the physical education banner, which is a traveling trophy that recognizes good attendance, effort, and behavior during P.E. She tries to get them to discuss their feelings about this, and gets them to plan how they intend to get the banner back from the competing sixth grade class.

She announces the transition to the math lesson, which involves a familiar routine. Pupils find the unit in which they are currently working, she picks a monitor for the day, pupils either work on practice problems, the problems in formative tests (which the monitor checks), take a summative test over the unit, or receive instruction from either Ms. Anderson or Mrs. Harvey, the learning disabilities teacher who acts as general aid during math time. Instruction is almost exclusively of the "here's how to do the algorithm to get the right answer" variety. There is little conceptual explanation given. In this respect, Ms. Anderson's teaching of math is much like that of Mr. Armstrong (and indeed, of most intermediate grade math curriculum and instruction in our experience). What is different here is that pupils are responsible to their teams to make progress through the units. Although the author of the Team-Assisted Instruction claims it is a cooperative learning program and pupils have been arranged in teams of deliberately mixed abilities, the brand of cooperation seems to be primarily the joint record-keeping of tests that members passed or failed (with remedial work in the form of additional problems) and the collective discourse of "What book are you in?" or, "How much have you finished?" One effect of the program in Ms. Anderson's class has been the substantial spread of progress through the curriculum, with some still in the Advanced Addition unit and others in the Pre-Algebra unit. Ms. Anderson worries about this as the test approaches, because "Some of them haven't even been exposed to fractions yet." Another effect of the pupils controlling the rate of progress through the units is that many do not go as far as their abilities might suggest. The only incentive to finish one unit is to progress to the next—to have more problems to work and more tests to complete. The alternative to doing more is to socialize, and observers often note that the ratio of math to socializing decreases as the hour goes on. Pupils take advantage of the fact that the teachers are occupied teaching groups in one or another corner of the room.

On this day in early April, Mrs. Harvey works with a group on the floor in the middle of the room, going over the method for finding the greatest common factor in fraction problems. Ms. Anderson, meanwhile, calls together the pre-Algebra group, who are fussy today because the topic is foreign to them. They have been trying to work individually on a problem set that has unknowns in it. They want to know what the n's, x's, and y's mean. Mrs. Anderson tries to calm them down. "This homework is very hard. We'll get started on it together so you don't flip out." She explains that the x's are like question marks. Ma asks, "Why don't they just use a number?" She says, "They can't use a number because then it wouldn't be a question mark." Ma persists: "Why don't they use a question mark?" She tells them that x signifies "an unknown," and also tells them that "the dot between two numbers means multiply." The pupils grumble and remain confused. She tells them not to get discouraged just because the program had them do homework problems before the instruction was to take place. "I don't want you to fail this part just because you haven't had it, so I'm helping you now. And then in the next lesson they'll tell you how to do it." She talks to them about the importance of algebra, about how it is necessary to know for advanced math courses and for careers like engineering. Ma sulks, "I'm not going to be an engineer." Ms. Anderson says, "You don't know what you're going to be. When I was in college the first time, I didn't think I'd be a teacher." The best math student in the class says, "I hate math," and Ms. Anderson replies, "You don't hate math. You just hate that this isn't being fair to you. Just hang in there."

There is a vexatious transition to lunch, with several pupils being sent back to their seats for poking others in line.

Back from lunch 40 minutes later (she had allowed them an extra 10 minutes because they had been working so hard this week and "this week is so weird"), she begins the work preparing them for the ITBS. "Okay, you guys. You remember I told you that a lot of this week we're going to be focusing on specific things that will be asked of you on the Iowa test. These are the things that are just expected of sixth graders to know." To the question, "We don't have to know everything, do we?" She replies, "It's a state law that says that everybody has to take the Iowa test." Several pupils call out their opinions of the test, one asking why they have to take it if it is Iowa's test, another complaining about it taking a whole week.

Ms. Anderson handles these complaints by reminding them to think beyond the test to the rewards and the lighter work load to follow: "But keep in mind that once all these tests are over with, we are going to have a fun day to just blow off steam and have a party. And then at the very end of the year you're going to have another fun thing, a party at Golf and Stuff. But this week we're going to set it up. Now I am expecting that all of you, while we are going over these things, that you are really paying attention and that you are really trying to do the best you can. If you have questions, this is the time to ask them. Okay? So you feel really confident when you guys are taking the test. You don't have to get a stomach ache. So you feel good when you finish. You think, 'Hey, I know this stuff.' And you can sit down and do well."

She hands out the photocopied pages from *Scoring High*. Other sixth grade classes have already used them, as some answer ovals are already dark. The pages she hands out are on capitalization. There are 28 items, and students are to darken circles in "answer rows along the bottom of the page. On the page there are directions ("Read each exercise and look for a capitalization mistake. In the answer rows, mark the answer space for the number of the line with a mistake. If you do not find a mistake, mark answer space 4"), two practice items, and two test-taking tips: "Only one line in each exercise has a capitalization mistake. As soon as you find the line with a mistake, mark the answer space and go on to the next exercise." Each of the 28 items has the same format, which is identical with the ITBS. There are three lines of text, each line marked with the option number, the fourth option being, "No mistakes," and the text comprising either one or two sentences. A facsimile for this format is as follows:

1. 1) Of all six of my uncles,
   2) I most enjoy going
   3) to uncle Luke's house.
   4) No mistakes

Ms. Anderson begins by asking a general question, "What kind of words do you capitalize?" The kids call out, "Names," "Beginning of a sentence," "Any word," "My name," "Phoenix." Directing their attention to the worksheet, she says, "Say this is the Iowa test. You've been given the test. What's the first thing you need to do when you're given the test?" In humorous spirit, the kids call out, "Nothing," "Rip it up," "Read the directions," "Guess all over the place." She patiently answers for them, "First you read the directions. Read the directions for this page, R" After R reads the directions, Ms. Anderson asks another pupil to read the first example and says, "Okay, first of all, what are we looking for, J?" J replies, "Things that should be capitalized but aren't?" Ms. Anderson responds, "Okay. Exactly. Things that are supposed to be capitalized but they're not. We're looking for the mistakes." She then reads each of the first three items, has them look for the line with the mistake in it, then asks one pupil to tell them what line they selected. There is much commotion as some call out and others dispute various answers. Then she asks them to work out the rest of the 28 items on their own. She points out the "tip" on the page: "You will only find one mistake for each problem. As soon as you find the line with a mistake, mark the answer and go on to the next one. You have to remember also that these tests are timed. You've got to work fast." Hearing the fearful "oohs's" from the class, she hastens to add, "That doesn't mean you guys won't have enough time. Okay? So you don't have to just skip through it and mark down any old thing. Read through it, check it. But if you find a mistake, do you have to go through and read the next four lines? No. Go on to the next one."

After two minutes, S announces that he is done, which means that he probably just filled in the ovals with a pattern. The class reacts incredulously and there is a joint discussion about how many items they have completed. Several students ask others for answers to some items. Ms. Anderson reminds them to do their own work. She circulates around the room and responds to questions; most replies are of the nature, "Sometimes you just have to trust yourself. Trust your instincts. You know these things."

Fifteen minutes later, she goes over each item, asks an individual pupil to tell what answer he or she selected, asks them, "How many of you picked number one." Then she tells them the correct answers. She asks them to explain the answers they picked, like why Grandma needs to be capitalized in item number three. Ma answers, "Because it's part of her name. They're talking about a specific grandma." Someone recognizes that a word in an item is capitalized when it should not be, and Ms. Anderson clarifies that some of the mistakes in capitalization are words the test-makers have capitalized that should not be.

Finishing this recital, she asks, "How many of you have 100 percent so far?" About two-thirds raise their hands. One girl says she guessed on all of them and got four correct. There is a discussion about capitalizing certain words like French braids or French toast, which Ms. Anderson resolves by sending someone to the dictionary. In one case, regarding the capitalization of God, she reminds them of their unit in mythology where they had discovered that when one is referring to several gods and goddesses, such words are not capitalized but a specific one, like Mars, is. The pupils readily call to mind the mythology unit, in which they all enthusiastically participated earlier in the year.

She introduces the next page, which covers punctuation, and assigns the items that they will work on by themselves. Noise rises, and she has to remind them several times to be quiet and work independently. They repeat the same drill as for the capitalization items. Several times she tries to encourage them to pick the answer that seems best, even if they can't state the reason why. "Count on yourselves and the knowledge you have gained in reading," is the message she dispenses.

While they are working on their own, she spends 10 minutes with S, showing with close physical proximity (her hand on his desk and her eyes looking directly into his) that she cares about him and believes in his intelligence. She has related to the observers that S has a disturbing home life and a family that has made him feel stupid and inadequate. These feelings carry over into his school work. She has to work with him individually to get him to do his lesson. In such situations, he demonstrates his abilities and knowledge. Tests like the ITBS bring out his tendencies to believe himself unable and to give up rather than meet the challenge. Ms. Anderson feels this happens to many pupils in Jackson's population. Like many other teachers, she wants to work on test preparation primarily to inoculate these pupils against the emotional paralysis in the face of the tests and the feelings of stupidity the tests seem to engender.

The entire lesson has taken two hours of the afternoon. What remains of the day consists of resting and compensating the pupils for working hard. Since some of them are out of class for band, those remaining visit or play "quietball" until time for dismissal.

3. Explaining Content

In the third type of preparation for the ITBS, the teachers explain the content and skills of test-specific material. This happens in two ways: (a) The teacher might interrupt ordinary instruction to add or explicate test-specific content, or (b) the teacher might devote class time to do the worksheets that the district had developed to prepare pupils for the ITBS.

Although Mrs. Samuels does some coaching of format (Type 4), she spends most of her time preparing her second graders in Hamilton by explaining the content and skills the ITBS covers. She uses the material that the district supplies.

Mrs. Samuels Prepares for the ITBS

> It is one week before the test. Mrs. Samuels' day goes like this. For the first five minutes before announcements, the children examine the tadpoles that hatched over the spring break. Mrs. Samuels complains that the arrival of the tadpoles, a part of the SCIS science program, never is timed quite right. This year the kids even missed out on the hatching. Mrs. Samuels thanks them for having their parents come in for conferences and talks about everyone starting anew. "This week we have to get to work because next week is the big test. All we have is this week to practice." The children are well aware of the big test, as Mrs. Samuels has referred to it all year in statements like, "You'll have to know that for the ITBS," or "When it comes time to take the ITBS, you'll have to listen closely and I won't be able to help you." She has worked on the district test preparation materials for about an hour a day for the five weeks before the test.
>
> After announcements over the P.A. system, at 8:15 a bell sounds, signaling the beginning of the reading lesson time for all the primary grades at Hamilton. Some children leave Mrs. Samuels' class (the middle-ability group of readers), and several come in from the other homerooms. Mrs. Samuels has two reading groups, one high (Suns) and one low (Cardinals). The Suns have already finished Reading Mastery II, and she has elected not to go on to the next level, opting instead for work in the district basal. As is typical for a reading lesson, she starts the Cardinals on seatwork, explaining the worksheets, and then assembles the Suns in the back of the room to do the basal lesson. Then she reverses the arrangement, the Cardinals working through the oral work of Reading Mastery II. The seatwork packet for the Cardinals involves work on consonant blends, determining which pictures on the sheet started the same way as "block." She explains how to sound out the words and reminds them, "This is something you will find on the ITBS. You will have to find pictures that begin the same. Some will be blends."
>
> When the Cardinals come to the reading group, she follows the presentation manual, but adds tips that pertain to the ITBS that Reading Mastery does not cover at this level, namely compound words and contractions. As individuals or in chorus, the pupils read the words from the chart: that, mad, sly, crooks, come, here, like. The children sound out the words they do not recognize. On the next page they work on making the "ea" sound, also "wh," "ar," and "ou." After reading the next list of words, Mrs. Samuels says, "One of the words in this column is a compound word." J offers, "Something." She says, "Two of the words in this column have endings. What is the base word or root word of buying?" A girl offers, "Buy." Another girl points out that there are actually three words that have endings, including "woods" that Mrs. Samuels had missed. She responded with enthusiasm and praise, "Good job! You're getting to know those compound words and base words and endings. You will need to know those on the ITBS next week and for your other work and on the CUES whenever that is." They read two more lists, identify the compound words and words with endings, and then read the day's portion of

the Reading Mastery story. In the Suns group, Mrs. Samuels makes special note of contractions, which the ITBS covers but Reading Mastery does not. In each case she explains the purpose of contractions and how to combine two words into one. The children seem competent with this content.

Reading period goes on until 10:30, after which the class spends about 10 minutes on the ATF discipline system, determining points they earned for conduct and effort, then go to recess. Back for math, they begin with the calendar activity, which is a regular part of Math Their Way. But unlike its use in ordinary instruction, Mrs. Samuels now uses the calendar to orient the pupils to the test. Use of the manipulatives has also disappeared from her class, as she has moved almost exclusively to the paper-and-pencil, "work the problem to its correct answer" kind of math instruction that is congruent with the tests. Referring to the calendar activity, she says, "You'll need to know this on the test next week. You will need to know how to read a calendar on the test next week. You all should do well on that because we've been doing that all year." She has them count out the number of school days they have completed, 133, and has them identify how many hundreds, tens, and ones are in that number. "That will also be on the test. See, you already have things you know. You just have to make connections."

She announces that they are going to do something new today and has them move close to the front of the room so she can demonstrate a page from the district *Math Objectives Review* (MOR) booklet. She shows them a large paper penny. "We talk about heads and tails. Lincoln is on the penny. Every coin has the year in which it was minted. What's on the back of the coin?" There is a little discussion about Lincoln, the Lincoln Memorial, and the like. She points out that there is information on the coin telling how much it is worth. She works with the nickel next and tells them it is worth five cents. She then talked about the picture of Jefferson and his home in Monticello. She asks, "Which would you rather have, a nickel or five cents?" When a boy answers the nickel, she asks why, and he replies, "It's worth more." She explains that they are worth the same amount. "This is the coin kids get confused with. What's a dime worth?" She talks about how a dime is worth more than a nickel even though it is smaller in size. She tells them that there are several ways in which other coins combined can equal the worth of a dime. This is to prepare them for the MOR worksheet, which is similar to some of the ITBS math items (Exhibit Four). She shows them a quarter, asks whose picture is on it and how much it is worth, and relates it to "a quarter of a dollar" or "one-fourth of a whole." She then pins some paper coins to the board in various combinations, and has them figure out the value of the total. She calls on individual pupils to give answers, and some are right and some are wrong. She asks them to explain how they got the answers they did. To count the money, she suggests that they should start with the coins that have the largest value. She puts up a dime, a nickel and three pennies and says, "There are two different ways of doing this," and "The thing that makes it hard is you have to switch gears in your mind." A boy calls out the right answer while she demonstrates counting out the numbers in order of their respective values. They go through several more of these. Then she starts with addition and subtraction with the coins, and suggests that they practice at home tonight with members of their families laying out different combinations of coins. Later, they would do some of the problems from the MOR booklet on their own.

For 15 minutes, Mrs. Samuels leads them through a drill on map reading. This is a worksheet the district developed and matches exactly both items on the ITBS and instructional worksheets of *Scoring High.* Among the children there is considerable misunderstanding. They grow restless as Mrs. Samuels' voice increases in volume and apparent tension. The children are not getting this; they lack the frame of reference that comes from familiarity with similar exercises.

After lunch, the pupils notice that they are not getting a chance to write today, and they are not happy about it. Journal and story writing and sharing, an important part of ordinary instruction in Mrs. Samuels' room has already diminished substantially and now nearly disappears until the tests are over.

Instead, they do another review sheet for the ITBS. This one is also over unfamiliar curricular territory, which reduces the exercise to coaching formats and test-wiseness rather than explaining content. "This is a practice for the way some things are going to be different from what we normally do. One thing that is very important. Listen to all directions. I'll read and you'll be filling in circles. I can only say it one time. So I'll read. Don't fill out something ahead of time. On the ITBS we won't be able to share answers. So don't call out. There is no way you will call anything out. Use a number two pencil, and don't mark them any harder than they are. Fill in the oval under the numeral—that's another word for number—that makes the number sentence true (Item number 3)." She takes them through these items, which are in a form completely different than their ordinary math work. She tries to piece together the vocabulary and prior knowledge that are necessary to answer the items but that have not been covered in the same form in class. She apologizes occasionally (e.g., "Granted that we've never had anything like this"). There are more incorrect than correct responses, and explaining and achieving consensus seem too burdensome. The kids are becoming understandably frustrated and restless, and Mrs. Samuels whispers to the observer behind tight jaws, "I hate this. We haven't done this, and this is what they'll miss on the test." She tries occasionally to reassure them. "We haven't even had multiplication, and most second graders haven't, so just don't worry about it." To an ESL student, she says that he won't have to take the test, so not to let the practice pages worry him.

This has been a tedious half hour. Afterwards, they go to P.E. and then to music. The last half hour of the day will be spent with the tadpoles, but by then the teachers and pupils are exhausted.

4. Coaching Format

The fourth type of preparation for the ITBS also involves worksheet drills, but emphasizes coaching in test formats rather than explaining or teaching new content that the test covers. What we mean by coaching is equivalent to the sentence, "If you come to situation X, do Y." Explaining is equivalent to the

Exhibit Four

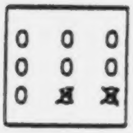
SECOND GRADE MOR REVIEW SHEET

I.8. Equations, Inequalities, and Number Sentences

1. 14 ◯ 6 = 8

=	+	-	x
○	○	●	○

2.



○ 7 + 2 = 9
○ (3x3) - 7 = 2
● (3x3) - 2 = 7
○ (3+4) + 2 = 9

3. 6 + (5-2) = □

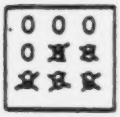
9	13	8	11
●	○	○	○

4.
○ 5 + 6 = 6 + 5
○ 4 + 1 = 5 + 0
● 7 - 4 = 4 - 7
○ 7 - 0 = 8 - 1

5. 8 + □ = 13

4	5	6	7
○	●	○	○

6.



○ 4 + 5 = 9
○ 9 - 4 = 5
● (3x3) - 5 = 4
○ (3x3) + 4 = 13

7.
○ 3 + 5 = 5 + 3
○ 2 + 1 = 3 + 0
○ 5 - 1 = 4 - 0
● 3 - 2 = 2 - 3

8. 15 ◯ 8. = 7

x	-	+	=
○	●	○	○

9. 9 - □ = 4

4	5	6	3
○	●	○	○

10. 3 + (8-2) = □

13	12	9	10.
○	○	●	○

Exhibit Four

SECOND GRADE MOR REVIEW SHEET

I.B. Equations, Inequalities, and Number Sentences

Note: Teacher reads each problem. Students only have answers.

1. Look at the number sentence in row 1. One of the symbols, when placed in the circle, will make this number sentence true. Fill in the oval under that symbol.
2. Look at the picture in number 2. Fill in the oval next to the number sentence that best tells what the picture shows.
3. In row 3, look at the number sentence. One of the numerals, when placed in the box, will make the number sentence true. Fill in the oval under that numeral.
4. Look at the four number sentences in row 4. One of the sentences is NOT true. Fill in the oval next to the one that is NOT true.
5. Look at the number sentence in row 5. One of the numerals, when placed in the box, will make this sentence true. Fill in the oval under that numeral.
6. Look at the picture in number 6. Fill in the oval next to the number sentence that best tells what the picture shows.
7. Look at the four number sentences in row 7. One of the sentences is NOT true. Fill in the oval next to the one that is NOT true.
8. Look at the number sentence in row 8. One of the symbols, when placed in the circle, will make the number sentence true. Fill in the oval under that symbol.
9. Look at the number sentence in row 9. One of the numerals, when placed in the box, will make this sentence true. Fill in the oval under that numeral.
10. In row 10, look at the number sentence. One of the numerals, when placed in the box, will make this number sentence true. Fill in the oval under that numeral.

Exhibit Four

SECOND GRADE MOR REVIEW SHEET

I.E. Decimals and Currency

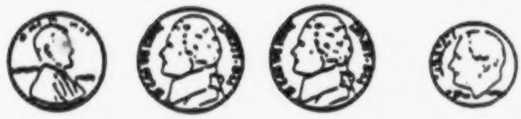
1. 45¢ $45 ¢45 405¢
● ○ ○ ○

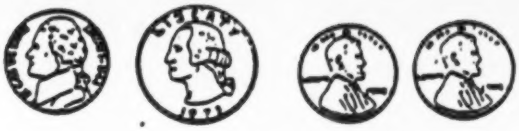
2. ¢70 $70 70¢ 07¢
○ ○ ● ○

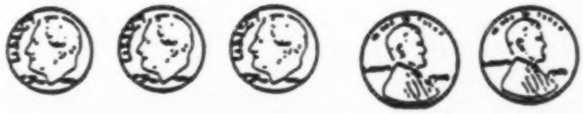
3. $33 303¢ ¢33 33¢
○ ○ ○ ●

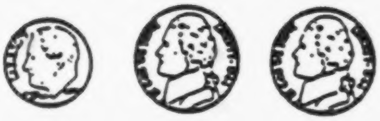
4. 27¢ 72¢ ¢72 702¢
○ ● ○ ○

5. $63 63¢ 603¢ 36¢
○ ● ○ ○

6.

16¢ 21¢ ¢21 17¢
○ ● ○ ○

7.

¢12 28¢ 32¢ 50¢
○ ○ ● ○

8.

23¢ 32¢ 17¢ 40¢
○ ● ○ ○

9.

30¢ ¢30 15¢ 20¢
○ ○ ○ ●

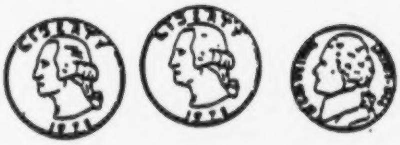
10.

15¢ 51¢ 55¢ 505¢
○ ○ ● ○

Exhibit Four

SECOND GRADE MOR REVIEW SHEET

I.E. Decimals and Currency

Note: Teacher reads each problem. Students only have answers.

1. Look at row 1. What is one way to write forty-five cents? Fill in the oval under a way to write forty-five cents.

2. Look at row 2. What is one way to write seventy cents? Fill in the oval under a way to write seventy cents.

3. Look at row 3. What is one way to write thirty-three cents? Fill in the oval under a way to write thirty-three cents.

4. Look at row 4. What is one way to write seventy-two cents? Fill in the oval under a way to write seventy-two cents.

5. Look at row 5. What is one way to write sixty-three cents? Fill in the oval under a way to write sixty-three cents.

6. Look at the pictures of money in number 6. Fill in the oval under the numeral that tells how much money there is.

7. Look at the pictures of money in number 7. Fill in the oval under the numeral that tells how much money there is.

8. Look at the pictures of money in number 8. Fill in the oval under the numeral that tells how much money there is.

9. Look at the pictures of money in number 9. Fill in the oval under the numeral that tells how much money there is.

10. Look at the pictures of money in number 10. Fill in the oval under the numeral that tells how much money there is.

sentence, "This is like this, because of Y." Although it is possible to use "explaining" about test formats ("When you are not sure of the correct answer, it is best to guess because test publishers employ formulas to correct for guessing"), such explanations are usually too technical for elementary school pupils. Therefore, material whose primary purpose is to increase test scores employ the coaching mode: "If you don't know the right answer, make a check mark, go on with the other questions, come back to the one you checked, and make your best guess."

Teachers at Jackson chose to use *Scoring High* because they believe these materials (compared to other methods of test preparation) effectively raise ITBS scores with the least cost to ordinary instruction. Not assuming that the ITBS is congruent with their programs or even with any defensible educational goal, they view test preparation cynically. They believe that the district administrators take for granted the validity of the ITBS as a measure of school effectiveness and would likely remove either the Jackson principal or the Whole Language Program if scores were unacceptably low. Whether any of this is in the strict sense true, it is these beliefs that govern their actions. Of course, there are exceptions, notably Ms. Anderson, who care little about the test scores and act accordingly.

Although teachers feel they must use *Scoring High*, they hate doing it. Knowing that they were transgressing their philosophy of education, they grieve constantly for the time it takes away from what they deem important, and for "the loss of the flow" of a pupil-centered classroom. In the following vignette, we illustrate everyday life at Jackson under the influence of test preparation.

Mrs. Orlando Prepares for the ITBS

> It is 7:45 on Monday morning in Mrs. Orlando's second grade class, the first day after spring recess, and the mood is merry but not "hyper." As usual, Mrs. Orlando has each table take roll for itself, a practice she believes promotes a sense of mutual responsibility and community. Only Tr is absent, and Mrs. Orlando tells them that he has moved away. Although they are sad that he is gone, they congratulate themselves on their perfect attendance. Mrs. Orlando asks how many would like to change their seats, "Just for a change." All but five vote to change. "Interesting," says Mrs. Orlando, "I used to be the kind of person who liked to stay in one place."
>
> On the P.A., Mrs. Mitchell announces the birthdays of the day, welcomes them back from break, leads them through the pledge, the moment of silence, and singing a new song. She reminds them of the lottery they will conduct to select a class that will represent them in a district special event.
>
> Mrs. Orlando calls them to the large group area for the calendar. Because it calls for a change of month, calendar work is more complicated than usual. She asks who has birthdays in April and tells them they can bring in treats if they like, to celebrate each. "I want you to try to figure out where I am going to put the 1 [first day of month]." "How many think I should start it here [she points, locates Friday, Saturday, then Wednesday]?" The majority are correct. "The first is a special day, do you know what it is?" Several join in a chorus of "April Fools Day." She asks if anyone played a trick or had a trick played on them for this day. Six children tell stories about it, and everyone listens carefully as each child narrates.

Mrs. Orlando asks, "Do you think April Fool's Day came about because of some superstition?" [They have been doing a project on superstitions.] "I'll bet that the reason why it started is written down somewhere in a book. We haven't been to the library in a while." Someone suggests looking in the encyclopedia. Mrs. Orlando: "The encyclopedia would be an excellent place to start. And we also have our wonderful librarian to help us."

The conversation about superstitions, tricks, and the calendar interweave. She asks them about the number of days in the month and the year. "Remember we talked about the number of days it takes for the earth to go around the sun?" The guesses are 4 days, 4 years, 29 days, 200 days, and she says, "Will everyone listen to this smart boy who knows the answer?"

"Remember we talked about spheres [she points to the board where there is a display.] It takes 365 and 1/4 days to travel around the sun. Remember when we talked about fractions." She shows them one-fourth and one-fourth, for example, equal one. She shows them again the trick of counting months by using the knuckles and spaces between the fingers. The mountains have 31 days and the valleys have 30, except for February. They had already seen this trick on a Sherrie Lewis videotape. They try this together and the kids agree, "That works." So they collectively decide that April has 30 days. Mrs. Orlando tells them she always remembers the poem that her mother taught her, "Thirty days hath..."

Someone says he thought the sun traveled around the earth, and Mrs. Orlando gives a two-minute lesson in the history of science: "We used to think that until an astronomer named Galileo discovered that..." This all goes on until 8:25, when she redirects them into writing. As usual, one activity blends into and overlaps another in Mrs. Orlando's class.

She says that they have "a lot to share this morning," and several students went to places during break that they would like to talk about. "I have something to share with you; it has to do with our study of animals....I have been noticing that there's one person in our class who is so anxious to write her story that she has been working on it all the time we have been doing calendar. T, would you stand up and tell us about it?" Someone says she shouldn't have been doing that, but Mrs. Orlando says she is willing to overlook that this time. T says she lost part of it. So Mrs. Orlando asks if anyone else has a story they have been working on, and N says yes but doesn't want to share it yet. Mrs. Orlando says, "All right, this is what we will be working on this week. We must work on our writings and finish those stories that you're interested in putting into the Young Author's contest. She tells the three or four students who want to read their stories now to wait until writing class.

Then she gets them into an activity about moving desks and emotional reactions surrounding such a move. She says that instead of an official move, they will have a trial move. Someone asks, "Is that a court thing?" "A trial seat. Trial comes from the word try," she replies. She asks what kinds of feelings they are having in their new places, and they jointly explore their reactions to the new seating arrangement. When the discussion runs dry, she sends them to their new seats to write in their journals.

At 9:02, kids quietly sharpen their pencils and get to work. Mrs. Orlando: "I'm going to give you 10 minutes to write, so make a choice about what you're writing about and get to work." After 15 minutes of writing, it is time for test preparation.

Mrs. Orlando: "Just leave your journals on your desks, because now it's time for us to be practicing for the Iowa Test that you're going to be having next week. You're going to be tested all week, so I need your attention while I talk to you for a minute, all eyes on me. Listen carefully. Next week it is very important that you all be here because that's the time that we're going to be doing the test . We'll be doing the testing usually in the morning, so if you're going to be going somewhere, like your doctor's appointment, B, or maybe the dentist or something like that, would you please tell your parents not to schedule the appointment for the morning. Try to wait until afternoon. You will be tested every day next week. We will have a very different schedule. And this week our schedule will be different also, because what we need to do is to do some more practicing for the test. So that when you get the test, it's not going to be something that you've never seen before and that you don't understand. Because that can kind of throw you off, so that even if you know some stuff, but you've never seen something before, you know you might go, 'What's this?'"

J interrupts, with a look of concern, "What is the test for, anyway?" Mrs. Orlando: "What is the test for? That's a really good question. Well, it seems to be that they just want to see how much maybe ...." N tries to complete the sentence, "How smart we are?" J chimes in with a possibility, "To see if the teacher is teaching us?" Mrs. Orlando: "Not how smart you are. Yeah, it's kind of like a check to see how the teacher is teaching, how well you're learning, and things like that. So that we can look at it and say, 'Maybe we should be doing something different next year. Maybe we should change the way we're doing something.'" A asks, "If someone got every answer right, could you like skip a grade?" J says no and Mrs. Orlando agrees: "No, and the test is not really designed so you can get every one right. Because the test has hard questions on there that you, you're not even expected to get right."

E: "What if you don't know what they mean? Like last year Mrs. Thomas didn't even know what it meant so we skipped that one."

Mrs. Orlando: "Sometimes you have to do that. Because tests can be confusing. You know, it's not that you don't know what's on the test and what they're asking you, but sometimes they can be confusing and they make a problem for us."

Mrs. Orlando: "So if that ever happens, you should stick up your hand real fast and say, 'I'm just kind of confused about this.' And you shouldn't just sit and go, 'Oh gosh, I don't know the answer to this,' and get real upset about it."

T: "You should ask the teacher."

Mrs. Orlando: "Yeah, or just make the very, very best choice that you can."

J: "Yeah they can only read it twice and then you have to just do it yourself."

Mrs. Orlando: "It's better to pick the best answer. It's not a good idea to just leave it blank and go on. It's better to pick an answer. Because that way, you might be right. But if you don't put anything down, can you be right?" [unison no.] "You could be very wrong. Remember that book that I told you about? Quite a while ago. Remember it was math and we worked some on the overhead projector with that? I'm going to give you a book of your own, just like that, and you're going to work on it, as much on your own as possible and we're going to do some work together in class."

A asks for a book to take home to practice for the test on. Mrs. Orlando says, "The best kind of practice you can do at home is to read every night and to write every night. And maybe to do some work with capitals and periods and stuff like that at home. So if you want to practice at home you should write something like a letter, because that's going to be on the test. And remember where the commas go and the capitals and all that stuff."

They are working on the *Math Objectives Review* (MOR) booklet that the district distributes. She says she is going to read the problems for them, "Because remember how I said that I would have to read some of the problems to you on the test?"

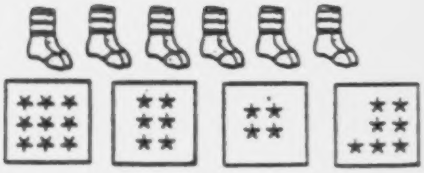
Mrs. Orlando: "Now this is what you have to do on the test. You have to listen very very carefully. And you will need to do your very best work. Are you ready?" They say yes. "Okay. This [Exhibit Five] says to me, 'Teacher reads each problem and students only have the answers.' You are to fill in the little circles underneath. Does everyone understand?" They say yes or yup, with quiet resignation or determination, but no apparent anxiety.

Mrs. Orlando: "Number 1. Look at the pairs of socks in row 1. Fill in the oval under the box that has as many STARS as there are PAIRS of socks." She waits about a minute and then repeats this sentence with the same emphasis, then waits another 30 seconds. "Okay. Number 2. Look at the number line in row 2. Now when I say number 2, I think I'd better caution you that when I say number 2, I'm about ready to read the second question. And what might happen if you still haven't finished number 1?" Several answer and she says, "Yeah. You just have to make up your mind and just pick one. Pick the one that you think looks the very best. Or I might say, why don't we do this, if I say 'Ready for number 2' and you're not quite ready, why don't you raise your hand and maybe I'll give you another couple of minutes to answer. How's that? Okay. Ready? Number 2. If you haven't picked your answer, what should you do?" "Raise your hand," they reply in unison. "Right. Raise your hand and that will be my signal and I'll go, 'Oh, I'd better wait a minute.' Number 2. Look at the number line in row 2. Fill in the oval under the numeral that tells where the arrow is pointing. Fill in the oval under the numeral that tells where the arrow is pointing....If you are

SECOND GRADE MOR REVIEW SHEET

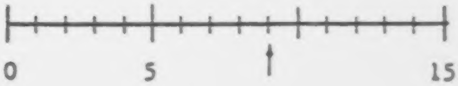
I.A. Numeration, Number Systems and Sets

1.



○ ○ ○ ○

2.



9 10 11 8

○ ○ ○ ○

3. 4 10 40 6

○ ○ ○ ○

4. 40 29 10 20

○ ○ ○ ○

5. 20 + 100 + 6 = ☐

21006 216 126 10026

○ ○ ○ ○

6. 4, 6, __, 10, 12

7 8 9 0

○ ○ ○ ○

7. 41 14 104 401

○ ○ ○ ○

8.
○ 3 + 5 = 5 + 3
○ 4 + 1 = 5 + 0
○ 6 - 2 = 2 - 6
○ 3 - 0 = 4 - 1

9. 3 4 5 6

○ ○ ○ ○

10. (3+2) + 5 = 3 + (☐ +5)

2 3 5 8

○ ○ ○ ○

confused about the question, what should you do?" She gives them a comic signal to use.

A asks what a numeral is and Mrs. Orlando says, "It's just a number." J asks, "What if the number [the correct answer option] isn't there?" Mrs. Orlando: "AH! What if the number isn't there?" Someone volunteers that they should mark the N, but there is no N on this problem. Mrs. Orlando: "But there is no N on this one, so it [correct answer] has to be there. Good thinking!....Is there any other confusion?"

"Number 3." N raises her hand and says, with a little discouraged voice, "I'm confused." Her problem is that she can't distinguish the numerals on the number line from the numbers that delineate the answer options, which are printed too close together. Mrs. Orlando goes over and shows her how to distinguish them.

"Number 3. See the numerals. What are numerals?" Together they shout, "Numbers!" "Fill in the oval that tells how many TENS are in 46. How many tens are in 46? So what are your choices? Are there 4 tens in 46 or 10 tens or 14 tens or 6 tens?" There is a one minute pause. T seems to be having trouble figuring it out. Someone whispers the answer to someone else.

"Number 4." Two hands go up. "Okay. A couple more sentences....Some of you are making this harder than it is. If you want to, take your chalkboards out. And what am I asking you to do? Write the number 46 on your chalkboard." She pauses and J asks if they have to use their chalkboards if they already know the answer and she says no. "Remember that I told you that you could use your chalkboards? Write the number 46 on it. Now how many tens are there in 46? Don't tell. How many tens are in 46? Maybe it will be easier if you see it. It would be easier for me." J says it's easy either way.

"Number 5. Okay. One problem I'm going to tell you about is rest rooms. What happens when you go to the rest room is that it really disrupts the testing next week. This is just practice this week, but if M [who had just asked to be excused] were to go out to the rest room we would either have to wait for her or do it at another time. And it's going to be very hard to make up these problems that she misses. So we'll have to give you a rest room break before we start the testing. It's VERY important that you go to the rest room when you get your break so that you don't have to go during the test."

Someone asks, "What about recess? Won't we get recess?" Mrs. Orlando assures them that they will. "And you'll get breaks and you'll get lunch and everything. All right, are we ready for number 4? Look at the numerals in row 4. Fill in the oval under the numeral that is 10 less than 30. If you want to put your chalkboard on your desk, " and she demonstrates writing 30 on her board and takes 10 away. "It's weird, isn't it? It's the way they are saying it. We don't talk this way a lot." Several kids have the "a ha" experience, but E says, "I don't get it." Although everyone in the class can correctly subtract 10 from 30 and get 20, they are unfamiliar with the wording of the item and lack a frame of reference for interpreting this language form and translating it into a form that they can correctly answer. Part of what Mrs. Orlando is doing is to provide them with workable frames of reference for attacking the items.

Mrs. Orlando warns J not to call out the answers or to say how easy it is. "We'll talk about it later, but right now I want to see how much you know without me helping you very much." They are practicing test-taking, not learning arithmetic.

"Ready? Number 5. Look at the number sentence in row 5. Fill in the oval under the numeral that makes the number sentence true." Some of the pupils look confused, again at the wording of the items. She repeats the instructions. "What is the number sentence? Twenty plus 100 plus 6 is equal to what?" Several little voices. "Don't tell!....Another way of thinking about it [the "number sentence"] is, 'what is the correct answer to it?' What would the correct answer be? What number makes the number sentence true?....Remember you can't write in the little box, you have to write on your boards....Number 6." Several ask her for more time. Some collaborate to get an answer. Some look puzzled. Some whisper that it is easy. When Mrs. Orlando tries to go ahead, Ja says, "I'm still thinking." She tells him, "You'll have to go a little bit faster." A asks J for an answer and he says, "You should be able to figure it out, A, if An and I can."

When they encounter items with the wording, "Which of these number sentences is not true," some of them think that it means, "how many of these sentences are true."

She reads number 5: "'Julie is fifth in line. How many people are ahead of her?'" E objects to the question: "I don't understand it because we don't know how many people are in line."

They work number 10 together. It is also unfamiliar territory:

$$(3 + 2) + 5 = 3 + (? + 5)$$

Several say, "I don't get it." E says that you have to do the parenthesis first.

Number 2 on the next page is one of those where nine circles are in a box, and two of them are crossed out. A complains that these are hard. K is on the wrong column of questions from the one Mrs. Orlando is reading.

Mrs. Orlando uses the tactic of saying, "Now what I would do..." to model strategies for attacking an item.

The kids are getting restless. Several yawn. Several lean their heads on their hands or desks. There is more whispering, commenting on the work of others, and who has finished or who doesn't know an answer. N falls out of her desk. At 10:00, Mrs. Orlando collects the booklets and they go to recess. Later, when she has a chance to grade these worksheets, she notes that each child missed several items. Most erred on the items with the crossed out circles, the ones with number sentences, and ones with parentheses. She doesn't know whether she can safely return these papers to them, meaning that they may get too discouraged by the number of wrong answers.

The thought that the difficulty of the materials may actually lower the confidence of the pupils occurs to many teachers who use them.

At 10:25 they are back and Mrs. Orlando invites them to the large group area for snacks and a story. The story is *Juma and the Magic Jinn* by Joy Anderson. She shows them the book and asks them to guess where the story takes place. They guess Asia, Africa, and Australia because of the cover illustration of a black person. "But there are black people here," she tries for more specificity. They name the turban and beads the character is wearing as suggestive of an exotic locale. She reads the story to them, pausing to show them the pictures. Her expression is wonderful, and she holds their attention with sidelong glances that maintain eye contact and emphasis on certain dialogue and words. She rarely interrupts her own reading so that they will not lose the sense of the story, although she once asked E what the word "script" means and asks someone to find Kenya on the map and the globe. They discuss the meaning of the mangrove pole and how they might figure out what it means from the story. The kids are utterly quiet and absorbed.

From the story, Mrs. Orlando leads them into an exercise with maps, which in turn will lead to a *Scoring High* preparation for the "visual materials" subtest. They each have world maps that can be written on with grease pencil. She has them put their finger on India, find the Indian Ocean and the Arabian Sea, and color in Kenya—all locations that figure into the story of Juma. She reminds them that there are five oceans in the world, and together they locate them on a flat map. She asks them if the Pacific is one ocean or two, and shows them how flat maps have to break up or distort, but they can see on the globe the relative locations of things. This continues until lunch at 11:30.

After lunch she directs them back to their world maps and has them locate north, south, east, and west, trace their finger from North America, starting on the Atlantic side, to Kenya. She tries to get them to see the connections between maps, globes, and spatial representations. She asks them to put away their world maps and get out the maps they made of the fairgrounds, the site of a recent field trip. She has them locate the entrance and the path they took by the Indian dancers, stage, petting zoo, and mineral building.

Mrs. Orlando: "Now I'm going to give you a map that is kind of like the map on the test next week." This is from the district materials, but is exactly like Lesson 20 in *Scoring High* [Exhibit Six]. They each have a copy, and she asks them to put their names on them. She tells them to take a few minutes and just look at the map. "This is a make-believe park, and now we are ready to answer some questions on this map. Which way is north?" Several say "up," or point to north. But there is confusion about the other directions; some know and some do not. She asks, "What are some things you see on this map?" Some hands go up. Some call out features of the map.

"What are the streets on the map?" There is a noisy response. "Read the questions together. Ready. You are entering Gigantic Park. What is the first thing you come to after you enter the park and go south? Use your finger.

Exhibit Six

NAME______________________________

DATE______________________________

SUBJECT__________________________

MAPS, CHARTS AND GRAPHS

ITBS format questions

For grade 2

Use MAP 1g

1. What is the first thing you come to after you enter Gigantic Park and go south?

○ Blue Lake
○ Duck Pond
○ Merry-Go-Round

2. In which part of the park is the picnic area?

○ Western
○ Southern
○ Northern

3. Which two things are closer together?

○ Library and Duck Pond
○ Playground and Picnic Area
○ McRonald's Farm and Blue Lake

4. Which is farthest from Blue Lake?

○ Duck Pond
○ Merry-Go-Round
○ Library

5. Which is directly east of Duck Pond?

○ Library
○ Merry-Go-Round
○ Swings

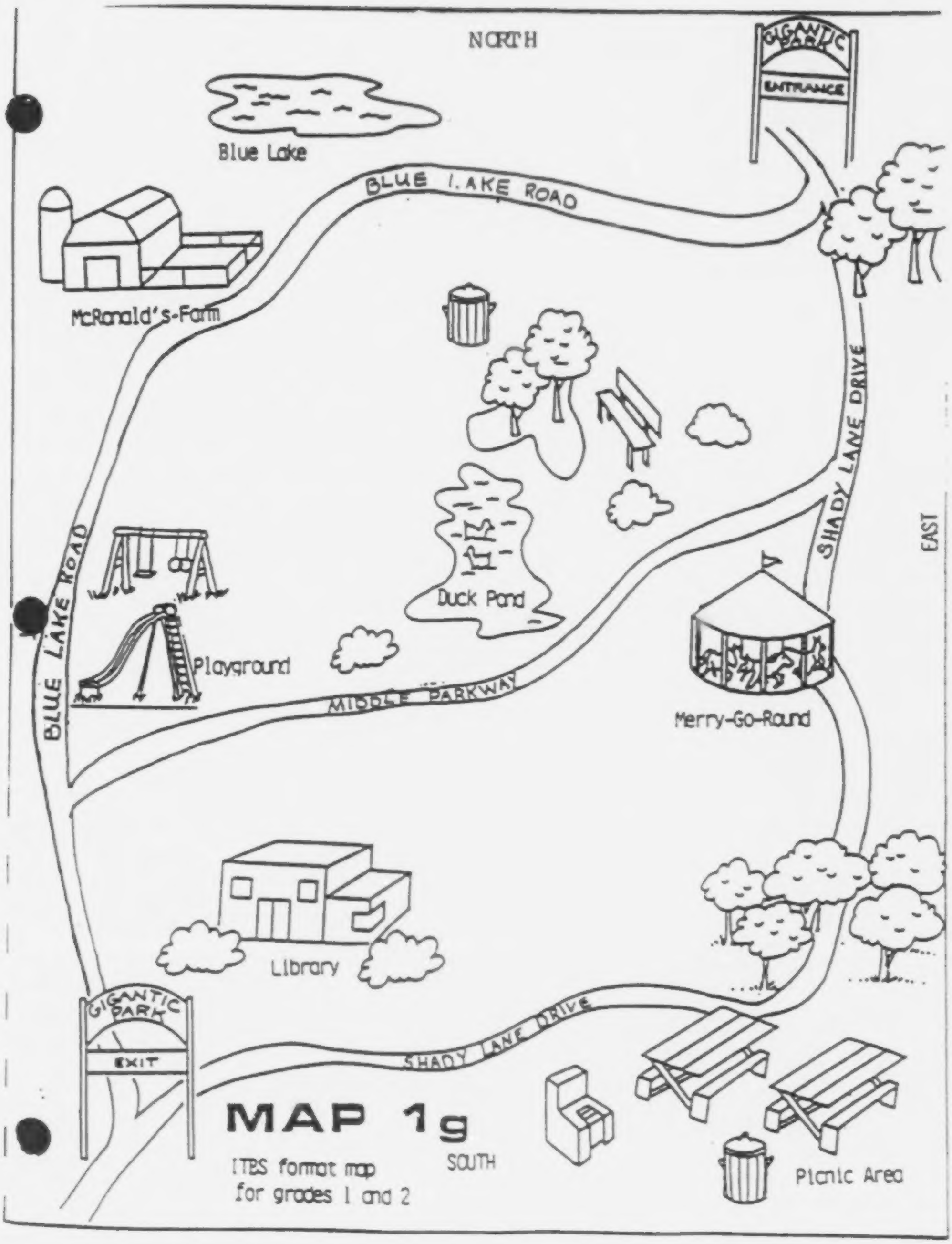
6. What road crosses the park?

○ Blue Lake Road
○ Shady Lane Drive
○ Middle Parkway

7. Which of these is on your right as you leave the park on Shady Lane Drive?

○ Picnic Area
○ Merry-Go-Round
○ Library

Exhibit Six


NORTH
GIGANTIC PARK
ENTRANCE
Blue Lake
BLUE LAKE ROAD
McRonald's-Farm
SHADY LANE DRIVE
EAST
BLUE LAKE ROAD
Duck Pond
Playground
MIDDLE PARKWAY
Merry-Go-Round
Library
GIGANTIC PARK
EXIT
SHADY LANE DRIVE
MAP 1g
ITBS format map
for grades 1 and 2
SOUTH
Picnic Area

It will be like your body. We're going to let our fingers do the walking." There are five different answers, including the merry-go-round. She turns her back to them so that her map and theirs will have similar orientations. She raises the map over her head and points with her fingers, following the specified route.

There is plenty of confusion about this; some of it is semantic (e.g., on the word "crosses"). When Mrs. Orlando explains the meaning of the words, the children can answer the questions.

"'Which place is farthest from Blue Lake?' You have three choices: the Duck Pond, the Merry-Go-Round, and the Library. Which place did you pick, An?" An says she hasn't picked yet, so another child answers "the library." Mrs. Orlando: "But I thought the picnic area is farthest." Several children call out that that wasn't one of the choices. Mrs. Orlando praises them lavishly, saying that they are so smart to pay such close attention to directions and not to get thrown off by this red herring.

As they go through other such items and orientations, such as in what direction would one go to move from point a to point b, the children grow restless, and she seems to be uneasy about the extent of incorrect answers and their inability to frame the questions in such a way that they can answer them. Some apparently are not secure in knowing their directions, but even several of those who do know them have trouble with the wording of these items. Children call out answers, many incorrect, or say that they are unhappy and do not understand. There is plenty of squirming. It is difficult to tell who is paying attention to this task. J is speaking out almost constantly. A is obviously bored. Mi and T are talking to each other. This class rarely does worksheets, so Mrs. Orlando has to stand over some of the kids to make sure they keep going and finish the paper. Several times she has to remind them that "your job isn't done until you fill in the oval."

When Mrs. Orlando is aware that many have failed to answer an item correctly, she says, "Some of you are right and some of you are wrong." When many are wrong, she takes them through the entire thing again. When they finally struggle through to the end of the worksheet she makes a face of exaggerated relief.

She gives them a short breather and then passes out Unit 1, Lesson 1, Vocabulary from *Scoring High*. She tells them that they will do the first one together, which provides a picture and three words, and they are to pick the word that tells what the picture shows. She calls on J to tell what he put down for the first one, which is a crude drawing of a woman with a bun standing at a door with a 2 on it, apparently knocking. He says, "Knock....This is kindergarten stuff." There are nine such items that the children work through on their own.

She also works the first one in Lesson 3 in Vocabulary, that is a completion exercise. "'A bright student is a ___________ person.' The three choices are friendly, smart, and light." M says petulantly, "What KIND OF bright? The LIGHT kind of bright?" Mrs. Orlando makes a face. This is one of those items

> where the children try to outsmart the test, read too much into it and are thus penalized.
>
> She tells them to do the paper by themselves. El says, "All RIGHT!" M immediately goes to Mrs. Orlando's desk for clarification. T goes up to ask, "What if I don't know the words?" Mrs. Orlando tells them to read all the words, "Don't just make a snap judgment after reading the first one. She says, "If you are stuck on a word, put your fingers over parts of the word and try to sound it out and figure out the word for yourself. I don't believe I can help you." Despite this admonition, children help each other—this is their typical mode of classroom learning—and many ask for her help. This lasts until the end of the day, a day that held almost nothing for the children except test preparation.

Metropolitan. Among teachers at Hamilton, there is very little special preparation for the Metropolitan Achievement Test, the purpose of which is to document the effects of Reading Mastery. Not only do teachers feel that extra preparation is unnecessary, but they feel that pupils need to recover from the ITBS and preparation for it.

Basic Skills Test. In the first two weeks of May, third through sixth graders take the district's Basic Skills Test. To prepare pupils for this test, teachers simply make sure that there are no glaring omissions in their coverage of the contents of the test, which they have known since the beginning of the year. Some teachers also repeat material already covered, but the intensity of the review and preparation is considerably less than their preparation for the ITBS.

Testing in Stage Five. From one to four weeks before the ITBS, teachers substantially reduce the time and energy they normally spend in ordinary instruction so that they can prepare their pupils for the test. They do this by intensive review of what they normally cover, perhaps altering the sequence of topics, explain or teach new content that they know the test covers but they ordinarily do not, coach pupils in test-taking skills and specialized formats, and try to promote pupils' sense of competence and self-confidence in the face of testing. The categories of test preparation represent differences in teachers' intentions. The worksheet activities teachers use, whether *Scoring High* or district materials, are virtual mirror images of the ITBS in format, content coverage, difficulty, complexity, and appearance. After Mehrens and Kaminski (1988), one can argue that practicing such worksheets is like practicing on alternate forms of the ITBS itself and having the teachers explain each answer option.

Test preparation for tests other than ITBS has similar qualities but differs in degree, in keeping with the relative power of the tests to cause shame or trigger district actions. Ordinary instruction diminishes, and time spent on untested material (e.g., writing, science, social studies, computer literacy) diminishes to near zero.

Stage Six: Testing and Preparing for the Next Test

As Arizona law prescribes, the second week of April is the week for testing children from first to eighth grades on the ITBS. Districts have some liberty to vary the order of subtests and to decide how many days of the week to use. Some districts pass these decisions on to the schools. There is little room for deviation on other things, for example, whether to administer the test at grade level or instructional level (the former) or what special education or ESL categories of pupil are exempt from the test. Exhibit Seven shows the specific instructions the district gives to teachers about standardized administration and handling of the test.

The tests take up about 90 minutes each day for 5 days. To most observers as well as the district administrators, this ought to leave the bulk of the day for ordinary instruction. However, the reality is quite the contrary. There are two general modes of activity for the remainder of the day—preparing for the next test and resting from the day's test. In this section we present two vignettes: days in the life of two classrooms, both in Hamilton. The two schools are not notably different from each other in this stage of the testing event.

Mr. Armstrong's Class Takes the ITBS

> This is Day One of ITBS in Mr. Armstrong's sixth grade class. It is 8:00, and the kids "escort" themselves in from the playground, passing the door with a humorous sign which asks for quiet during the tests. R comes in muttering, "Tests, tests, tests." Hearing this, Mr. Armstrong observes, "Every year you sense a little bit more tension when the tests come up. They're more important, I guess."
>
> A student new to the class, C, already faces the wall on Step One. As is typical of most days in this class, there is a pattern of noiseless communication, all meaningful glances and gestures and the occasional noiseless whisper and passed note. N is killing time putting little teddy bear figures on the tips of her pencils. J, E, C, and Ju are reading library books. F is examining a drawing he did. Mr. Armstrong asks, "T not here? Ci not here? Would someone put their chairs down, please?" Absences during test week create problems of makeups and worries about the class average (not having T's score included would likely lower Mr. Armstrong's average).
>
> The playing of the Marine hymn over the P.A. dramatically breaks the quiet. F smirks at E What associations do they have? Dr. Thorne begins announcements with date and time and a moment of silence, over which Mr. Armstrong talks to Mc about what excuse she has for her absence.

Dr. Thorne: "I want to remind all students that this week we will be taking the Iowa Test of Basic Skills. I would like to remind students that it is important for you to do your very best on the test today and through the week. Let's make sure that Hamilton students come out as well as we possibly can do and do our very best. I would like to congratulate a young man who made 100 percent on his spelling test last week. A special congratulations go out to Z. He has really been doing quite well lately, and I want to congratulate Z for his good work." More smirks; everyone knows what a goof-off Z is. "A note from Mr. Marquez's sixth grade class. As I understand it, all students in this room have memorized the 53 propositions—prepositions, in language class. And Mr. Marquez's class is challenging other sixth grade classes to memorize those prepositions and maybe have some kind of contest

Exhibit Seven

ITBS TESTING FOR APRIL 11-15, 1988

You should have the following materials:

Directions for Administering
Answer sheets (Grades 3-6)
Instructions for Teachers
Class Printout
ESL Exempt List
Teacher Questionnaire

1. On Monday, April 11th, you may pick up the test booklets at the office.

2. For 1st and 2nd grade teachers, there is no test form answer sheets for Levels 7 and 8.

3. Answer Sheets:

   a. Bubble in student number starting at #5 and going to the right.
   b. Bubble in the necessary information for date of birth and sex.
   c. DO NOT bubble "Group or Other Information".
   d. Bubble number of continuous years this pupil has been enrolled in district in the column marked CLASS D, under "Special Class Reports."
   e. Use legal names and not nicknames.
   f. Fill out the Pupil Variable Information on the back. On #3, mark only if child is ESL. LEPs are indicated on your student list.
   g. There is no need to alphabetize the test booklets or answer sheets.

4. Keep test materials under lock and key when not in use. NO MATERIALS are to be taken OFF THE SCHOOL PREMISES.

5. Timelines at grades one and two are guidelines except for the Math Computation, which is exactly 4 minutes. All number lines must be removed from the desks and walls. Slates may be used instead of scratch paper, but it may dramatically slow them down since the computation is timed.

6. Practice Tests: Do each practice test before each subtest. Do not give the entire practice test on Monday. Practice tests are for Grades 1-5 only.

7. Please double check all your test booklets and answer sheets for correct information on the front and back. Also check for stray marks. This is VERY IMPORTANT!

8. When turning materials in:

   a. Make sure all tracking strips are running the same direction.
   b. Place the Teacher Questionnaire on the top of the stack, matching the tracking strip.
   c. Use the paper band provided and fasten with tape. NO PAPER CLIPS PLEAS
   d. For grades 3-6 only, please bundle test booklets with string in sets of 25.
   e. Turn in all extra materials.
   f. Practice tests do not have to be turned back into the office.

9. ALL TESTS AND TEST MATERIALS ARE TO BE TURNED IN BY FRIDAY, APRIL 15th BY 3:30 P.M.!!!!!!

0. Make-up tests will be discussed at this meeting as to what will be appropriate.

after this week is over. Best of luck to all students taking ITBS, and let's make this a really fine week. Thank you."

Return to quiet. Mr. Armstrong says they will not have to worry about filling in the name and other identifying material on the test answer sheet grid. "I've done all that for you. I'd love to fill in the answers for you, but the state won't let me do that." F asks what time they will take the test, and Armstrong says during reading period today through Thursday. "I'm assuming that we will follow this method. We will start with vocabulary. That should run about 30 minutes. I need to look this over while you're gone to P.E. And then the second one is reading comprehension and that is about a 42 minute test. So we'll have some time left over. That should be a PRETTY easy way of doing it. Then tomorrow there will be spelling, capitalization, punctuation, and usage. OH, I'M GLAD we reviewed THAT! That covers a whole big area. I don't know how we're going to do on that. The vocabulary I'm not sure of. The reading, I think we're going to do well. I just think there will be no problem there. Spelling, umh, umh (his hand wavers back and forth in a gesture that shows he is not sure or that their performance could go either way). Punctuation, we ought to do well. You ought to do excellently on math. I'll be looking the test over while you're out to P.E., but I bet you we've covered just about everything that's in there. That one doesn't come up until Thursday, so we'll still have a little more time for review [many sighs]....What's the formula for finding the area of a circle?"

J says "Area equals pi r squared."

Armstrong: "Right. How do you find the circumference?" When A and K in turn respond incorrectly, confusing area with circumference, Mr. Armstrong reacts. "That's the area. That's what I'm worried about. I was afraid of that. We need to review." He provides formulas for area and circumference, and they write them again in their notebooks.

They flee this mini-prep session for P.E. class. During the next hour, Mr. Armstrong looks over the test for the first time this year. The other sixth grade teachers make brief visits, and one says, "We have met the enemy and they are us," quoting Pogo. Armstrong laments not having a scoring key so that he can tell how the kids performed. He made up one for the Basic Skills Test, then he computed the percentages so he would know in advance how the class performed and could check the accuracy of the printout when he received it in late May. On the BST-Placement, Hamilton got 89 percent of the items correct in reading and 86 percent in math. In reading, that was the highest score of any of the feeder schools. Surprised with this outcome, he notes: "And that was even adding in the scores of C and R." R needs to be in the transition class, according to his assessment, but the school refuses to make the placement until R 's records arrive.

It is unusually quiet on campus today. For once, the sounds of cricket snappers and word-calling in Reading Mastery are silent. The doors to classrooms are almost all closed, and the signs are up that warn against making noise while testing is in progress. One of the other sixth grade teachers reminds Armstrong to keep his kids quiet during room change, because the "Third graders have already started testing." The rhythm has changed. The day is gorgeous, the last of the springtime before the descent into the desert summer heat. All the coolers are turned on, partly for white noise, partly to cool.

Mrs. D, the sixth grade transition teacher, comes in to ask Mr. Armstrong about his schedule for shop class last week. Since she was reviewing for ITBS, she had not yet started her group on the schedule. Armstrong explains that half the kids go for half the grading period to shop class, leaving social studies. Then the groups reverse. "Everything comes out of social studies. It turns social studies into hamburger. That's what we take away from when we review."

Mr. Armstrong explains that, "To give them an edge on the math test," he skipped problem solving and went to geometry, going very fast in the hope that maybe a few will get the idea and "give them an edge." He says they have gone through decimals and fractions, "which is where we should be for this time of year. After the test, we'll wipe our brows and then go back to problem-solving....But I'm not so sure that's not good teaching. Because when I go through it really fast, they concentrate and when I go through it the second time, they're familiar with it."

The kids come in, sweaty from P.E., reporting their performance in various field events. Ms and A complain that they don't feel good and want to go home. A kid from another class comes in to report A and Ms for kicking and pushing him on playground during P.E. They compromise on a subtraction of five points from their ATF total for the period. No one seems to take this altercation very seriously.

Reviewing for the math test occupies the entire math period, specifically repeating material from the geometry lesson, which they covered the day before and recovered later in the afternoon.

He asks what a circumference is. Su calls it "the perimeter of a circle." Armstrong: "I'm not sure that's the proper way to say it, but it makes sense." P stumbles over the formula for circumference, "three and one-seventh times diameter." Armstrong: "Let's make sure we know what diameter means." They mumble and finally one answers correctly. He has them figure out the circumference of a circle with diameter of 35 cm. They convert to improper fractions and work through the steps. Su and Sh get the correct numerical answer, and he reminds them to express the result in centimeters. They then work with a problem using diameter expressed in decimals instead of fractions. Armstrong asks N for the formula for the area of a circle. "A equals umm, umm." Five hands in the air. P gives an answer which is the area of a triangle instead of a circle. Someone else gives the formula for the area of a rectangle. Armstrong says, "Let's work some problems. It's so darn simple." He points out some problems in their book and gives them some time to work each one. He prompts them with, "What is the first thing that we do?" "What is the next thing that we do? We're going to work this out, step by step, like we always do."

On one problem that gives them particular trouble, Mr. Armstrong says, "Oh, my! We've got some work to do on this." Looking past the individual difficulties, however, this turns out to be a moderately successful exercise, about two-thirds of the items answered correctly by someone, but there are many errors just the same. Although there might be some doubt about the extent of their understanding, there can be little doubt that they can do the calculation once they determine which algorithm to use.

At 9:35 he tells them to close their books: "The learning happens when you get into the position of dredging it up from your memory and then checking to see if you're correct. And you go through that process until you can retrieve it properly." One can tell that Mr. Armstrong is getting worried about this. The tension level rises with each incorrect response. His solution is to work more problems, "step by step." When N misremembers the formula for the area of a circle or the meaning and value of the term pi, he remonstrates, "You're bright; if you intended to learn, you would know it by now. But I know what you're doing, you're just waiting for it to go by. You have got to know the little details." Even after he writes the formula on the board (from Su's response), N can't read it, saying "Area times pi squared." "Look up this way, don't look at your paper, let's try it again." When she finally reads it correctly, her classmates applaud, but she still confuses perimeter and pi, ultimately calls it "area within" the circle. Mr. Armstrong says he "Hope[s] there will be something that sticks. We were better off on Friday. We need to spend more time on this." He screams in semi-mock frustration, when An can't answer the last question he asks, the meaning of the area. "Let's pray for a better day." Mr. Armstrong's class will later achieve an exceptionally high score on the ITBS, despite this day's stumbling and bumbling. Either the test preparation did the trick or the anxiety was undue.

The review session continues until 9:55, when it is time for the testing session to begin. He first tells them to put sharpened pencils on their desks so he can see that they are there. There is great rustling and bustling, which annoys Mr. Armstrong. He cruises the room to make sure the pencils are number two lead and sharpened, and sharpens some himself. He puts up the testing sign on the door after reading the humorous messages written on it by a previous class. He asks them to clear off their desks. They are quiet and waiting patiently. E tries to sneak a peek at the test on the teacher's desk. Su asks to pass out tests, but Armstrong does it himself. Several kids have books waiting for them in case they finish early. Mr. Armstrong hands out the answer sheets, on which he has already bubbled in the identification information (the teachers had gotten together the previous Friday for a "bubble party"). F gets a laugh when he points out that Mr. Armstrong had incorrectly classified him as a female. Armstrong has incorrectly identified him as female. Mr. Armstrong thanks them for reading quietly while this is going on, but asks N whether Stephen King is her parents' choice for her reading material. "Seeing the books some of you read, it's hard for me to imagine you sitting on your mother's knee reading these books. But as you have said, I am not of this world."

They receive the green form of Level 12 of ITBS. He reads the instructions out loud, "While you read them to yourselves." He reads without expression, rapidly, not emphasizing the important parts of these complex passages. "'You are now going to begin taking the Iowa Test of Basic Skills. It is very important that you do your best on these tests, otherwise they will not really show how well you can do reading, language, mathematics. Remember that we want to know whether you need more help in learning these subjects, so make the test a true picture of you by doing the best that you can on each one."

Everyone but C and E reads along, apparently concentrating on the directions. E is sleepily rubbing his eyes and yawning. C fingers some object. A stares out in space. Mr. Armstrong warns that some people will take their tests while sitting against the wall, "if that's what it takes to keep the right order."

The first test is the vocabulary test. "Which of the four words mean the same as the word in heavy type? Fill in the oval of the word that means the same." They go over several practice exercises together and listen to more instructions on how to use the separate answer sheet—make heavy marks, stay within the ovals, erase completely, only use the number two pencil, keep your place on the answer sheet, make only one mark per item—a familiar drill by now. "If you do not know the answer to an exercise, leave it and come back to it if you have time." He gives them no directions about guessing. He tells them that they have 15 minutes to take the test. E raises his hand to ask if they can fold the answer sheet (but the directions disallow it). K asks to have the blower on; they take a vote and decide it should be on. Mr. Armstrong improvises more directions: "Pace yourselves. Don't waste a lot of time on one question that you don't know the answer to." He asks if there are questions, but says "You've taken these tests so often, you're bubble experts by third grade."

When Mr. Armstrong says "Go!" (10:12) C startles, jerks his head and his arm toward the paper. K notices that Mc is already on page two, within one minute of beginning. Ms looks at Su's answer sheet; N looks at Ja's, Sh looks at F's. The pupils are concentrating and look intense, with no particular stress evident. No one seems to be bubbling at random, although Mc's rapid progress is suspicious. Armstrong is working at his desk. The door is open.

Mc is done at 10:18, checks over her test very quickly, and opens a book to read. Mr. Armstrong is now on his feet, circulating. "I don't want anyone reading a book...until 10:23. Until that time I want your nose in the test. Check it again." Mc rechecks, and makes two changes. Armstrong is back at his desk. J finishes, realizes that it is after 10:23, and opens her book to read. N finishes next. By quitting time, everyone seems to have a mark in each item and to have had time to check his work, although not everyone does so.

Armstrong goes right on to the next test, reading comprehension. He tells them that they have 42 minutes to get it done and admonishes them, "Start pouring on the steam to get it done." Sh points out to Mr. Armstrong that her test booklet has marks on it; Mr. Armstrong gets her a new one. At the beginning of the test, N has her hand in the air, but he doesn't see it, so she eventually puts it down and goes on. Armstrong: "No library books until 10:55."

E's lips move as he reads passages. No one is sneaking peeks. It looks like intense concentration. Mc coughs a lot. F pulls out several Kleenexes from his bag, P bends down to meet the desk with his chin, R frequently breaks attention and stretches, Ms puts her sunglasses on and begins playing with her fingers. F's foot beats staccato in the air above the floor, occasionally brushing it, N is cold, rubs her arms and puts on a sweatshirt. Ju shakes her hand, indicating it is tired. E keeps slogging. Nothing hints he isn't trying, despite Mr. Armstrong's view that he rarely tries on school work.

At 10:53, Armstrong cruises the room. K and Ju notice him behind them but don't look around. Armstrong looks at R's answer sheet, does a double take, moves away and sighs. R failed to finish the first test and it looks like he probably will not finish the second one. "He's going to draw the whole thing down."

At 11:00, Ms raises her hand but gives up because Mr. Armstrong doesn't notice. N has her book open. Armstrong asks J if he feels he did a good job. "Are you satisfied? Don't you need to go over it?" J is a source of vexation because he has qualified for Project Potential but rarely puts forth much effort on these tests and scores below grade level.

At 11:03, Mr. Armstrong tells the class, "You need to make an assessment now of how much you have left." He encourages them to work faster and finish if possible. But he doesn't say anything about guessing or random bubbling of those for which they lack the time to answer in the normal way. Other than the quoted instructions, Armstrong gives no messages encouraging the pupils to try hard.

By the end, most have finished, but few have gone back to recheck. Most read quietly. He calls time, tells them to read until lunch, goes over their answer sheets looking for stray marks, and locks the test booklets in a cabinet.

In the afternoon, he goes over another review of the same math material they covered this morning. This takes about 50 minutes. Then he conducts a review of the maps and graphs test they will take later in the week. Since the entire test booklet is distributed to the teachers on Monday, they have the opportunity to focus their reviews from Monday afternoon until Thursday on the specific materials covered on the test. Mr. Armstrong mentioned that he "lucked out" by diverting the schedule of math lessons. The regular sequence in math called for instruction in metrics. Instead he re-reviewed and stressed computation in decimals and fractions. The test has more items in the latter, he says. But now he must do some crash preparation on geometry, because he sees that the test covers it. He also plans to review the computation of decimals and fractions again before Thursday's math test.

Signalling over the P.A., the secretary asks him to please send a girl down to greet a new student. All hands are up to volunteer. Armstrong: "Just in time for the test. I hope she's a genius."

Armstrong: "Okay. Listen. We have about five minutes, and I've got to get the new student lined up. Really, I think you're in good shape. Everything that we have gone over today we've gone over in the past (Ms interjects, "40 times"). The thing that I'm concerned about is, not that you haven't been exposed to the material and that you know it; the thing that I'm so concerned about whether you have the patience to read it thoughtfully and answer it to the best of your ability. That's my concern. I ask you to do that, and you'll come out with a decent grade. You cannot get a decent grade skimming through it to get through it as fast as you possibly can. There's no way you're going to make a decent grade on it. So I'm going to let you have these five minutes free to read quietly if you like. Or you can take out your language book and skim through it. I'm not going to hold my breath until I see anyone do that, but that would be a rather encouraging sign."

Armstrong greets the new student, Ml, whom most kids already know. She was at this school before, then transferred to Jackson, now is back here. Armstrong: "I forget, Ml, are you a mental giant? We were just getting ready to take the ITBS test." Ml smiles, shrugs, shakes her head, taking the remark in the humorous and ironic spirit that he intended. Armstrong: "Well, Ml, are you happy to be back?"

Mrs. Samuels' Class Takes the ITBS

This is a third of five test days for Mrs. Samuels' second grade class. It is 8:00 a.m., and the kids are still filtering in, as they tend to do in this class. Mrs. Samuels remarks with some surprise and pleasure, "Ph, you made it! And homework too! Good heavens. Terrific!" Ph is one of the top-performing students, but almost always late, sometimes absent, and always diffident. L is crying because he doesn't have his homework. "It's my mom's fault," he sobs.

There is a general, pleasant discussion about getting their money in for some event. Mrs. Samuels: "J is still not here. Hopefully, K will still make it. Makeups are such fun," she says with an ironic tone. Mrs. Samuels directs the students to move their desks, as T explains, "We have to do this for the ITBS." After the move, the desks are about three feet apart.

The P.A. announcements come on. Dr. Thorne asks them to observe a moment of silence, which they do in a routine way (nobody seems to be praying or meditating or whatever this state-mandated moment is supposed to promote). He reminds teachers in intermediate grades to send one student from each grade to judge the young authors' stories. He congratulates those rooms that have "made their day" at high rates and singles out one boy, who normally misbehaves, for having made two double days in a row. He announces that the primary student council will hold a meeting and that the library will be closed from 12:00 to 2:00 for ITBS makeups. He encourages everyone to keep up the good work on the ITBS.

Mrs. Samuels passes out the test booklets and pencils she has already sharpened. She tells the students to put away all their crayons and coloring pages and not to open the test until she tells them to.

When they clear their desks, Mrs. Samuels signals the opening with, "Okay. The practice test will cover everything that we will do today. She seems to be reading, and her voice is artificially cheerful and formal. "At the bottom of page 5 is L-1, Spelling." She goes over the practice exercises, which they work through together. This community working through is much like normal routine for this class, where someone knows each answer and contributes the knowledge to the collectivity, so that it appears that everyone knows it. When they finish, the aid collects the practice tests and hands out the test. Tn is not here. As an ESL student with one year or less in school, he does not have to take the test, so he is spending the morning with the ESL teacher.

Mrs. Samuels reads, "Open your booklets to page 15....This is a spelling test. I will say three words and then use them in a sentence. The three words are printed in a row in your test booklet. You are to fill in the little oval under the one that is spelled wrong. Look at the sample exercise, marked S1. The three words are ran, top, and hill. 'I ran to the top of the hill.' Which word is wrong? (Unison response: 'hill', which is printed "hil" in the booklet). Yes, the word 'hill' is wrong. So the oval under the word hill has been filled in. We will do the rest of the exercises the same way." According to directions, Mrs. Samuels must read the item number, pronounce the three words distinctly,.read the sentence slightly emphasizing the three words, and then pause for time. There are 29 of these.

During the reading of the directions, De is rolling his pencil around his desk and Mrs. Samuels says, "Relax, Cr." R is visibly nervous, with lots of atypical body movements and rolling his chair around. Mo grumbles about the testing taking all week. Mrs. Samuels tells her, "We're almost through." Obviously angry, Ra starts to cry. With a broad gesture, she sweeps her pencil box off her desk and onto the floor, puts her head down on her desk and cries. The aid goes over to her to pick up the box and comfort her. She complains that Su (her frequent rival and sometimes best friend) is looking at her test from a distance of six feet. Now Ra makes a big production of putting her glasses on, still casting angry sidelong glances at Su. It is difficult to imagine how she could be listening to directions or paying attention to the opening parts of this test during this tirade. She rolls her test booklet up into a cylinder and puts her face down almost to the page, perhaps to ensure that Su will not be able to copy.

At one point, Mrs. Samuels tells T to sit up. Several times children raise their hands, and Mrs. Samuels repeats the prescribed sentence. Otherwise, she seemed to follow the directions exactly, not overemphasizing the misspelled words.

At 8:22, (10 minutes late), K comes in. Mrs. Samuels directs her to go to her desk where her test booklet is waiting for her and start working; she can catch up with the class as they go along.

At 8:25, Mrs. Samuels reminds Ra to put her glasses on, and Ra does it with angry emphasis. A few of minutes later she makes another big production out of erasing a mistake, all exaggerated and angry gestures.

The items complete, Mrs. Samuels says, "Okay. Good job. Stretch a little bit." Most of the kids stand up. Several complain about the cooler being off or on.

After about four minutes, Mrs. Samuels says, with staged cheerfulness, "Okay, turn to page 16. Language-2, Capitalization. On this page are several short stories and a letter. You are to find the words that need capital letters. There are very light ovals under the first letter of some of the words. When you find a word that should begin with a capital letter, fill in the oval under the first letter of that word. Look at the example at the top of the page: 'i saw ann's new puppy.' The ovals have been filled in under the word i and the letter a in the word ann because both words should begin with capital letters. The oval under the p in puppy has not been filled in because puppy should not begin with a capital letter. Do the rest of the exercises in the same way. Be sure to make your marks good and dark. Stop when you get the stop sign at the end of the page." Interestingly, she adds to the printed directions the following: "Read each sentence, as best as you can, fill in the ovals that need capital letters. You may begin." She sees that several are on the wrong spot, and she says again, "the bottom of page 16."
A facsimile[5] of the first paragraph looks like this:

The book store is having

a sale of used books in november.

the manager has some copies of

peter pan for half price.

There are ovals between very narrow parallel lines under the b in book, the s in store, the s in sale, the n in november, the t in the, the m in manager, the p in peter, and the p in pan.

There are three other paragraphs, more difficult than the facsimile, that require correct capitalization of street names, names of odd cities and monuments, ethnic groups, and appellations.

As there are no time limits on this test, the students can proceed at a pace that is comfortable for them. Ra seems to be more on task with this section. Several hands go up at different times, and Mrs. Samuels or the aid go to the kids and try to help. The usual question is about what a particular word is, but Mrs. Samuels tells them she is not allowed to tell them a word and they must try to sound it out and do the best they can. She does not tell them to go back over when they are through to check their work, although several finish in five minutes.

---

[5] To avoid violating security of items and copyright protection, we created item facsimiles which we then submitted to judges to verify that the facsimiles and the original items or exercises are consistent with each other in content, format, and level of difficulty.

Mrs. Samuels whispers to the aid, "This is the part of the test that makes me crazy because there are proper names that are off the wall and names of cities they have never heard of and something like Pan American Airlines—what do they know about that?" In her opinion, this section of the test measures reading rather than language, because there is so much reading they have to do on their own.

When they are all done, she has them stretch up again and stretch down to touch their toes. They seem relaxed. When they ask to take a break, she says, "We'll take a break in a few minutes, let's do one more part first." There are many groans and moans at this. Mrs. Samuels: "Come on you guys, you're doing great. This is the first complaining you've done this week." More groans. Mrs. Samuels: "Okay, sit back down. Settle back in." Several people say they're tired. Both Su and Ra look very unhappy.

With forced cheerfulness she again reads from the manual, "Look at the top of page 17, ladies and gentlemen. This is a test on the use of punctuation marks. It will show how well you can use periods, question marks, and other punctuation. These are the punctuation marks you will need." As prescribed, she puts the marks on the board:

period .	comma ,
question mark ?	exclamation point !
apostrophe '	quotation marks " "

She reads on: "In the box at the top of the first column, you are to decide where periods belong—Is everybody on the right box?—Notice that there are very light ovals under some places in each sentence. You are only looking for periods in this part. When you think that there should be a period, fill in the oval under that place. Now look at the first line. It says, 'Mr Daniels said I should go home'. Two periods are needed, one after 'Mr' and one at the very end, after 'home.' Notice that the two ovals under these places have been filled in. A period is not needed after Sims, so this oval was left blank. Now go ahead and fill in the ovals under the spaces where period belong. Do not mark places where any other punctuation belongs, just periods. Be sure to make good dark marks. When you come to the stop sign at the end of this box, wait for directions. Ready. Begin."

There are six separate boxes, one each for periods, question marks, apostrophes, exclamation points, and quotation marks. The box for commas resembles this facsimile of a personal letter:

1944 Main St

Boulder Colorado 80302

February 20 1985

Dear Beth

Last week we had

a Valentine's Day party I

got candy hearts cookies and

paper valentines

Sincerely

Bobby

In the above, there are ovals between the 4 and the Main, after the St (note the distraction of the missing period), between the city and state, after Beth, after party (another distraction), after hearts, after cookies, after paper, after Valentine's, and after Sincerely.

They progress, punctuation mark by punctuation mark, through the test section. As she moves around the room, she notices that K is working on the wrong box and corrects her. She tries to encourage Ra to do at least one box.

After giving directions on the apostrophe box, she adds, "If you don't know, take your best guess. READ the paragraph." There are several murmurs from the kids when she says that this section is on the apostrophe. She complains to the aid, "We don't teach apostrophes except in contractions. Then they over-generalize and automatically put in an apostrophe after 's' because that is what they have learned."

After this paragraph, Mrs. Samuels says cheerily, but with an edge to her voice. "Okay. Hang in there folks. Next column. Okay. We know this. We've worked on this." She circulates. R has his hand up, and she looks at his page, saying, "You've got it. You know what you're doing, R." Her confidence seems warranted, as she gives the same message to several others. But to Si, she says, "Commas only, Si. Commas only."

When they finish, she says, "Okay. Two more paragraphs and then we can take a break. Here comes a short one." The kids moan and sigh and groan. They say they're too hot or too cold or they need a break. Mrs. Samuels and the aid exchange sympathetic looks and say to each other that the kids must be getting tired. Their tone seems to imply, "Who wouldn't be?"

On the exclamation point paragraph, Su protests, "I can't do this." Mrs. Samuels responds by reiterating the directions and encouragement, "Just read it and put quotation marks in where they are needed. You can do it." More moans greet her introduction to the section on quotation marks.

Even when they finish this section, she is afraid to give them a break to get a drink for fear of disturbing the other classes still testing. She lets them quietly go look at the tadpoles or talk among themselves.

The children move around and try to regroup, but some are genuinely unhappy. Da whines pitifully: "I'm tired." Su pleads with Mrs. Samuels, grabbing her arm and pulling on it, "Do we have to do anymore?" Mrs. Samuels: "But Su, you're doing a great job!" Su, pouting, "But it's sooo boring, a whole week!" The aid pats her on the back and says, "We're almost done." Their comments are reminiscent of a nurse holding on to a young patient during minor surgery, knowing it hurts, but only being able to say, "It will pass." The kids cheer up a little when Mrs. Samuels tells them she can read the sentences to them in the next section on usage, whereas they had to read the previous ones themselves.

"Come on, guys. Get your effort back together. Don't desert the ship now." Mrs. Samuels reads the directions: "This is a test on the use of words. In each box there are three sentences. One of the sentences is better than the other two. You are to look for the best sentence. Then mark the oval in front of it. Read the exercise marked S-1 silently while I read it aloud." A facsimile of the example is, Eddy goed to a party. He had a good time. It was the bestest party ever. "Which is the best sentence? Yes, the second sentence is best. Notice that the little oval in front of the second sentence has been filled in to show that this sentence is best. We will do the rest of the exercises the same way."

She reads the first item, of which the following is a facsimile. Them boys is playing ball. They hitted a home run. They play every day. "Fill in the oval for the sentences that sounds right." There are 27 of these, in a similar form.

Although the pupils work diligently, some are obviously tired or bored. Ra balances her chin on her desk with her chair pushed far back. Cr fills in answer ovals without looking at the test questions. Mrs. Samuels points out to the aid why it is that the pupils are so perplexed. Consider this facsimile item. Did you see the circus? My brother seed it Saturday. He don't no when he can take me along. "The kids hear it correctly, but read it 'no' instead of 'know." In other items, they can trust what they hear. In this one, there is confusion between the senses. "That makes this a spelling test, not a usage test." She claims that "with these kids, I have to correct it all the time in reading. In common speech they use phrases such as hid hisself. Even when they see the print 'himself,' they read aloud 'hisself.' Or 'I seen some pictures.' This is the way their parents at home talk. And we haven't had time to work with them to change it yet." But even these kids giggle at some of the sentences (e.g., "I like hamburgers more meself").

There is a whole chorus of oohs and aahs and sighs after the last one. The aid picks up the pencils and tests. Afterwards she and Mrs. Samuels will spend several hours cleaning up stray marks and darkening the ovals that the children failed to darken enough for the machine to read.

Since they can't go out to recess before the prescribed time (again, the taboo against disturbing the testing of other classes), she tells them that they may walk quietly around the room and then do "a coloring page." There is an immediate recovery of good spirits, except that Su and Ra's rivalry shifts fields. Now Ra is angry that Su is using markers rather than crayons to color the page. Mrs. Samuels works to get K caught up on the items she missed, although the testing conditions are certainly different, as there is noise and activity in the room, and Mrs. Samuels can look over K's paper as she works. As soon as it is safe, she leads them out to the playground for recess.

Mrs. Samuels spends the time critiquing the tests they have taken. She complains that some of the sentences are conceptually and logically too difficult for the kids, such as the one, "Was your mother ever a baby?" "How many of the first graders ever think about whether their mother was ever a baby? They don't have the concept. One of ours yesterday, the kids, including two of my better readers, raised their hands and said they just didn't understand what this is talking about: 'Is a building that is far away always larger than one that is close?' What does that mean? They could read it, they just didn't have the concept. So there. Are you testing reading or are you testing concepts that they may not logically have? And the spelling words just don't fit. The directions are 'which one sounds wrong?' And yes, the words are spelled for them, but if they're told to be listening for sounds. And some of them for the teacher are hard to read grammatically incorrect. The first year I gave it, I had a lot of trouble with that." Some of them just fill in the ovals and some of the better readers work too fast and make careless errors. Unlike last year, Mrs. Samuels has not had anyone crying. At least there wasn't as much crying this time. "Ra 's brother in first grade has been crying every day and throwing the book on the floor, like she did last year." Tomorrow they will do math, and Friday the "maps." Previously, they had done the test in four days, but "the powers that be" decided they would stretch it over five days, not leaving any time for makeups for anyone absent on Friday. But even if someone missed school on Friday, the district could still compute a composite score with the reading, language, and math scores, "which they pay more attention to anyway."

She points out one of the questions from yesterday "that really threw them." The *Phoenix Gazette* later reproduced this item, which the teacher reads to the pupils. "An orange, a cherry, and a watermelon are in front of you. If a cherry weighs more than an orange but less than a watermelon, mark the oval under the cherry. If an orange weighs more than a cherry but less than a watermelon, mark the oval underneath the orange. If a watermelon weighs more than an orange but less than a cherry, mark the oval under the watermelon. You should have only one answer marked." All she got on that one, she says, is perplexed expressions. "One of my pet peeves is all the farm questions," which might be appropriate for Iowa pupils, where they make up the test, but not for city children. She objects strongly to the last paragraph on the reading test being the longest, for psychological reasons. The kids are not familiar with names of states, cities, and books on the spelling and capitalization tests (e.g., Los Angeles Raiders). Nor does their curriculum cover quotation marks in second grade, much less apostrophes. It strains Mrs. Samuels to force them to work as individuals on the test, since they do their usual class work by helping each other and seeking help from her.

After recess, they do the calendar, emphasizing the pattern of pink butterflies and yellow umbrellas on the calendar, and counting by 10s to 140 (the number of school days already passed). But unlike most days, she truncates the Math Their Way activity in favor of review for the ITBS math test they will take tomorrow. She assigns them five "review pages" from the

Macmillan text series, which the textbook authors designed to prepare pupils for the ITBS. The district refers to these as the Macmillan correlatives. "This reviews everything we've done in math so far. We will be taking the ITBS math test tomorrow."

Although the children should be working the problems at their desk, they seem to have had enough work for one day. Uncharacteristically hostile, they run around the room, shouting, teasing and hitting one another. "You sound like a bunch of squabbling, whining preschoolers," she says.

"Listen to directions. Yesterday we had a problem with this. A problem like this is NOT subtraction with regrouping."

She writes on board:

$$\begin{array}{r} 15 \\ -8 \\ \hline \end{array}$$

"A lot of you got confused on this yesterday. Watch. I saw several people doing this. They crossed out the one and put the five here as if they were borrowing. You don't have to do that, folks. Just use your base facts. You already know that 15 minus 8 is 7. You got so lost on regrouping you forgot how to go back to basic facts. If this were 65 take away 8, then yes, we would have to borrow from here so that this WOULD be 15. This already is just 15. We don't borrow. Do you see the difference?" Some say yes. She shows how to count backward from 15, if they don't know the fact by heart, or count forward from eight, using their fingers.

As she continues going over the kinds of problem on the review sheets, she notes that some involve addition of three numbers. In such problems, "you may write them." On another page that has horizontal problems, she reminds them to "go ahead and write them up and down." She notes that K "is not listening to directions, as usual."

"This reviews everything we've done in math so far. Tomorrow we do the math ITBS. And some of it will be this format, with filling in the circles and some of it will just be doing the answers....I DO expect you to do it by yourselves. Think carefully about whether it is asking you to add or subtract. So get your brains in gear."

She reminds them of the format of items on the page and how it resembles that of the ITBS in that some options are "N" for no answer given. The kids have trouble getting started, and there is the usual moving about and getting communal answers, though she pays it no heed. "There really aren't that many, Ra, so pick one and get started." "You've got it, Sweetie (to Si). Keep going." To Tn, the ESL student exempt from the test: "I wish you were doing the math ITBS test with us, Tn, because you do great in math. But I would have to read it to you."

Gradually the kids begin to work more in earnest. It has been some time since they have done simple subtraction without regrouping and addition without carrying, so their "circuits have overloaded" and they try to make the process more complex than it really is. So she has decided to go back to simple problems so they "would be refreshed on math facts." She says that constant cramming for the test made this dropping back necessary. Because she had used Math Their Way most of the year, the class had fallen behind in the regular math book. Then, to catch up and emphasize regrouping (which the test covers), she had skipped over the units on counting time and money. Then last week she had to cram time and money counting, which most of them "got real fast, though two or three of them can't get the idea of counting money at all" (e.g., converting a nickel to a 5). [See Exhibit Four.]

As the kids complete their pages, they bring them to her desk for her to check. Those that are incorrect must be redone. Other pupils come to her when they are not able to work the problems.

Over the P.A., the secretary informs them that a new pupil will be joining the class. The student is Na, who had been at Hamilton last year. She was asked to repeat first grade, then moved to Mexico, and is now back. With more than a trace of wistfulness, Mrs. Samuels wonders whether she will be required to take the ITBS, either the remaining test or make up the full battery. It later turns out that Na falls into the ESL exclusion rule, which is a good thing, because she displays little English.

Now Na comes in and the entire class greets her with genuine warmth, with the possible exception of R, who says, "I thought we were going to get a GOOD-looking girl." Cr swats him and makes him be quiet. Mrs. Samuels determines, with some relief, that Na fails to respond to directions in English about how to do the math page, and only responds when Mrs. Samuels gives her instructions in Spanish. Mrs. Samuels tries to enlist Cl to translate for Na, but Cl claims (falsely) not to be able to do so. Seven children gather around Na's desk. Mo and some others remember Na from before and are particularly warm and friendly. They all try to help her count in Spanish, including Tn, to everyone's great amusement. Su and Ra compete for Na's attention and demand that her desk be between theirs.

So ends the morning. During the afternoon, Mrs. Samuels will let the children recover, read a book to them, and pass out treats as a reward for their good efforts on the test.

No single issue is so salient for teachers at this time of the school year as the deleterious effects that taking the ITBS has on primary grade pupils. Teachers feel the tests injure the pupils' psychological well-being and sense of themselves as competent learners. Almost every teacher reports anecdotes about children crying, wetting their pants, fighting, calling themselves stupid, stabbing themselves with pencils—on and on. The principals agree, one of them calling the negative effects "undeniable." Administrators at the district central office do deny them, believing that teachers project their negative attitudes about testing to the children and thus bear responsibility for any harm that ensues.

Direct observation of the two second grade classes revealed few instances of problematic effects on pupils. One of Mrs. Samuels' pupils was angry enough to throw her pencil box and scream at her friend. Anger, frustration, and fatigue were contagious in the classroom that week, causing the teacher to take measures to settle the pupils down. Whether the tests hurt the children's psyche is not likely to be settled by participant observation, however. Because our observations were spread so thin over testing occasions and classes, we could neither confirm nor disconfirm teachers' assertions of widespread impact. On the other side, we were close enough to a number of teachers to evaluate whether their claims were merely self-deceiving, self-interested, or likely true. Our access to their more private thoughts and feelings, as well as the pervasiveness of the reports, leaves us little room to doubt them.

For the intermediate grade teachers, the job is more than anything else to keep the pupils focused on the task. Few teachers claim that intermediate grade pupils suffer harm as a result of testing. Either the pupils are so accustomed to testing by this time in their career or they care little about tests or achievement generally, thus rendering the test harmless to the students, according to the teachers. The atmosphere in the class during testing is businesslike, although teachers observe many of their pupils "dropping dots"—filling in the answer sheets with interesting patterns. Teachers seem to recognize this behavior when their most accomplished pupils take 20 minutes and their least accomplished take 5.

Despite their aversion to tests and their beliefs that tests fail to measure achievement accurately, teachers administer the tests with precision—strictly by the book. Their posture toward the administration of the test is stoic, communicating no negative feelings about the test to the pupils. They are positive, encourage the pupils to try hard and do their best, and soothe them when they show signs of anxiety and fatigue. They often promise treats, breaks, and other rewards to keep the pupils from giving up. Even when reading an item they consider hopelessly difficult or ambiguous, the most they do is roll their eyes and slog through it, assuring the pupils, as Mrs. Samuel does, that "It is almost over."

Although every teacher claims secondhand knowledge of cheating—telling one's worst students to stay home the week of the test, prompting pupils on correct answers, providing extra time, erasing incorrect answers and replacing them later with correct ones—they deny doing it themselves. They say it would be unprofessional to do such things, at least under current circumstances. However, some said that if the stakes changed, and somehow their pupils' scores were to affect their job or position on the career ladder, that they would then find a way to make sure the scores were high. During our observations of testing, we saw no instances of teachers' violating standard procedures of test administration, although no teacher is likely to stretch the rules in the presence of even a trusted observer.

Certainly there are opportunities to cheat if one were so inclined. By state rules, on Monday the principal distributes the test booklet containing all tests they will administer during the week. Teachers could readily look ahead to tests they must give later in the week and "teach the test" itself—do repetitive drills on vocabulary or spelling words, for example. Engaging in such practice is apt to be effective[6] in raising scores. But if scores are too high, they will likely attract scrutiny by test administrators at the districts. In any case, we genuinely doubt that teachers at Hamilton and Jackson engage in such practices.

---

[6] According to Shepard (1989), if a third grade teacher remembers only one item from the vocabulary test and teaches it to the students so they can memorize the correct answer, that would result in someone at the 49th percentile raising his or her score to the 54th percentile. A class average can be raised five percentile ranks by learning two items.

More tests. According to some of the teachers, the injurious effects of testing on pupils occur because of the sheer number of tests the pupils must take. For the sixth graders, the CUES and curriculum-embedded unit tests are ongoing, the Study Skills Test is administered twice, there is the ITBS and two sets of Basic Skills Tests, one of which is administered in March as a placement test for seventh grade. Students in Hamilton also take the Metropolitan Achievement Test in reading. Tests piled on tests, and from the pupils' perspective, all cover more or less the same content, skills, and follow the same format. Only from the organizational perspective are the tests different, for they each serve a different organizational function. Teachers comply and the students endure.

All teachers denounce the drudgery and redundancy for the pupils and object to the necessity of encouraging effort on the pupils' part. They note, as well, the sheer amount of time taken away from ordinary instruction by the need to prepare for, administer, and recover from each test. On the other side, they recognize that familiarity with the content and format of one test likely raises scores of another. As one teacher at Hamilton remarked, the high scores the intermediate grades attained had as much to do with the number of tests they took and the similarity among them as with the efficacy of the curriculum and teaching.

Testing at Stage Six. If you happen to be the pupil, taking the standardized achievement test puts you face-to-face with these demands: read or listen to words and sentences that may be beyond your comprehension; select answers to questions someone else has decided are correct; work alone and don't peek; go as fast as you can; no matter what, keep trying hard, even when you are weary or hopelessly perplexed or when something outside school may be bothering you; keep quiet; get over the hurdle. If you are the teacher, the role demands of testing are these: read directions someone else wrote; impose time limits someone else has determined; don't provide answers or extra time lest you render the test invalid; keep the children quiet; do whatever you can to keep the children trying hard; don't frame questions for them by interpreting the meanings of words they might not understand; clean up the answer sheets so machines can score them; act as a professional; hope for the best.

Looking at the tests themselves one can see that many items are ambiguous, open to different interpretations, demand performances that children cannot readily meet or programs do not cover. The tests are so long and there are so many that inevitably fatigue and tedium and possibly loss of confidence, anger, and frustration result. The content that tests cover and the formats in which items are written make the testing more fair to some curricular programs and less fair to others.

In real time, tests consume only a fraction of the school day. Yet teachers feel pressure to review material that the next day's test covers. Otherwise, they spend the time on "R and R," rewarding their pupils for trying hard and making up to them for enduring an unpleasant experience. Precious little ordinary instruction goes on.

Stage Seven: Resting/Reorganizing School

The weeks immediately following the ITBS are the prototypic anticlimax. According to the district administrators, the teachers should already have returned to ordinary instruction and district Scope and Sequence. After all, a CUES reporting date approaches, and the BST in a month's time, and there remains nearly a fifth of the academic year. At Hamilton there is also the current levels in Reading and Spelling Mastery and the other textbooks to complete.

For few teachers does this image match reality. For most, including the four focal teachers of this study, the weeks after the ITBS are a period of resting and recuperating from the test. Teachers feel guilty that they subjected their charges to excessive anxiety and effort, not to mention the possible psychological injury the tests inflicted. They find ways of, if not exactly slacking off, at least failing to press ahead with vigor on those parts of the curriculum as yet uncovered. There are popcorn parties, school talent programs, field trips, longer-than-usual lunch and recess breaks.

Some teachers try to recover what they had to give up in favor of preparing for the test. Mr. Armstrong cycles back over the units in the math text that he had skipped. The pace is anything but frantic, and he later admits that the class never got around to completing the text. He reinstitutes writing, the subject closest to his heart, and the pupils respond with essays, stories, narrative descriptions, and poems that please themselves and their teacher. Soon after the ITBS, the class completes Reading Mastery VI. Afterward, he uses the reading period for community reading of some books of C.S. Lewis, having the pupils keep a journal of their reactions to plot line and characters.

There is a brief flurry of preparation for the BST. He makes sure that he teaches those portions of the science and social studies textbook that the test covers, although these texts also are not completed by year's end. He prepares them no further for the Metropolitan or the BST in reading and math, because, "they have been tested to death" on those subjects. Besides, the BST merely rehashes CUES and Study Skills. On the Social Studies BST, for example, only half of the items test substantive content (such as early civilizations, Greek, Roman, African, feudal, and Renaissance civilizations). The other half of the items test skills of reading maps, charts, graphs, and time lines, skills that were taught and tested in every intermediate grade and in reading, math, study skills, science, and social studies. Why review them again, he wonders?

Mrs. Samuels' class also enjoys a reinstatement of writing, clearly the favorite activity for her pupils. They quickly revive their community. Their previous pattern of collaborating on seatwork and obtaining a great deal of help from Mrs. Samuels had been suppressed during the test-preparations and test-taking. The pace of academic work is noticeably slower. The Suns have completed Reading Mastery II. Although the principal would like them to go on to Level III, Mrs. Samuels decides to work a little in the basal and do some literature studies. The Cardinals proceed with Level II, and may not complete it this year.

Ms. Anderson's class also does a little, almost desultory preparing for the BST, but mostly the class rejuvenates. She decides to spend some time on units involving current events. She sympathizes with the amount of effort they have expended on the various tests and compensates them with lots of play and some unfocused academic activities.

Mrs. Orlando's class is the closest to a return to ordinary instruction as she defines it. The pupils return to literature study and the study of units like the earlier ones on magic and superstition. They attempt to restore the sense of community and still have some fun. Even here, there is some kind of need to revitalize after the trials of the past six weeks.

As the school year grinds to a close, external tests play a peripheral role. Next year's placements must be made, but this year's test scores will not be available until too late, not until June. At Jackson, all the children who have spent this year in transition first grade will be placed next year in regular first grade. But Mrs. Mitchell decides to discontinue the transition program for next year. For the most part, a child's current teacher, Mrs. Mitchell, and potential teachers of the subsequent grade negotiate children's placements for next year. Later, Mrs. Mitchell will consult the ITBS score distribution as a crude indicator, to make sure that next year's classes are roughly heterogeneous with respect to prior achievement.

For sixth graders at both schools, junior high teachers and counselors use the BST results along with sixth grade teachers' recommendations to make placements into classes stratified by ability (e.g., into five levels of reading classes and three levels of math and science, or into programs that resemble Hamilton's transition classes).

At Hamilton, the school gets reorganized much as it was organized in August, with a series of TAP and grade-level meetings. In them, teachers, specialists, and administrators look at year-old ITBS scores, current course work and CUES results, and listen to the testimony of teachers. The options are retention in grade, placement in transition class from regular class, movement from transition to regular class, evaluation for possible placement in full-time special education programs, or combinations of placements and services. The process of TAP closely resembles TAPs earlier in the year and will not be illustrated in this section. For the bulk of the pupils, teachers' and administrators' conversations about pupil's current progress through Reading Mastery, language, and math determine their next year's placements.

Testing in Stage Seven. Freed from the demands of testing, the teachers use this time to restore their own priorities. But less energy is available to pursue them vigorously. Between the time they take the tests and the time the testing company reports the results, schools must reorganize for the subsequent year. The schools use the same mechanisms they used at the beginning of the year, but only year-old test results are available to help them.

Stage Eight: Reacting to Test Scores

Because they perceive that the test scores are largely out of their hands and are not sure what to expect, teachers hold their collective breath when the scores finally become available, about the first week of June. Table 3 contains the test results from the two schools on all the ITBS subtests in grade equivalent scores and growth. Most of the scores of both schools are near grade placement but lower than the district averages. In the district as a whole, there is a relationship between test scores and social composition of the student bodies.

Some of the participants provide their personal reactions to the scores of their pupils and those of their schools. Ms. Anderson says:

> I'm surprised at myself. All this year I've been saying how inaccurate those tests are. We get back the ITBS, and you know what? They are darn good. I've been thinking a lot of this year that I haven't been a very good teacher and wondering how much they've learned. We get back the ITBS and then I felt, "Well, maybe I haven't done so badly. But, wait a minute? Just because the test says so?" Isn't it funny?

Mrs. Orlando's initial reaction is that the test scores "weren't so bad." Before the scores came back, Mrs. Orlando made up a ranked list, not of the true achievement of her pupils but how she predicted they would score. When she compares her rankings with those of the actual scores, she sees few discrepancies. "The tests tell me about what I expected. What this all means is that we don't need to go through all that—upsetting the schedule, practicing, all that time taking the test. We know it already." She contemplates the three children whose scores differed from her rankings. She puzzles over one girl whose scores went down from last year's scores and wonders about the girl's passivity and lack of initiative. One boy fails to perform up to his potential in either classroom work or the test. Another boy, whom the district has labeled learning disabled nevertheless scored nearly at grade level. Although she recognizes he made amazing progress in reading this year, "I don't know how he could have scored so high."

As for Mrs. Mitchell, she reports being "not concerned" over the low scores in second grade spelling and the small (less than a year) growth from first to second. She attributes the former to the curriculum they follow at Jackson in primary reading and language. She thinks the low gain might have occurred because last year's first grade teachers engaged in too much test preparation, achieved high scores, thus leaving little room for improvement from first to second grade. She notes that these scores refute the board's charge: "The board has said the scores have gone down since I've been here, but these scores don't show that."

At Hamilton, Mr. Armstrong has this to say:

> I guess the kids came out very well on both the Basic Skills and the ITBS. And if I remember right, Dr. Thorne said the class just showed more of an increase than any other, if I'm understanding what he said right. I never did go back and read the scores. But he was really very delighted....But, they're not to be believed. (Laughs). I mean, it's just my opinion, but they're not to be believed. When I have children who are checking out at seventh grade, the eighth month or eighth grade, that's not to be believed....And I've been fortunate in the reading class, because I've had a super reading class. And they are, they're bright kids. And I don't have to worry about teaching them how to read, how to decode, that stuff. We just talk about concepts, and that's exciting. But I don't think, when I see two years' growth, I just don't believe it. It's really hard to tell. You don't know what to believe. Because the Reading Mastery—I really am a Reading Mastery person. I believe in the program. I don't like the script, and that type thing. But I can't deny the

grades. Everyone has made 90s. If they are not making in the 90s there is something really wrong. So it's designed for really high success. And the ITBS must be designed for really high success, maybe that's why it sells, maybe that's why districts buy it. Well, if I were making a test, and I were profit-motivated, I'd probably make a test that would make the receiver look as good as possible. The tests seem to be all skewing way up, way up high. I would love to feel like I was doing that great a job.

I didn't really study the printout. Dr. Thorne was showing it to me in the office. He was just elated because, everyone looked good. It doesn't stand out in my mind. I guess that tells you how I feel about what the tests mean, to me personally, they don't mean anything. Just doesn't mean a thing. But I think some of the reason why they're so high is because we've reviewed and reviewed and reviewed and reviewed. And that has to be some of the key, to have it fixed. And everything that we had studied in our book, we stopped in the simple geometry part of the book—we never did finish the book. We stopped right at the polygons and shapes and boxes and such, and I took the Basic Skills Test, and that's exactly, exactly where that test stops, at the area of a circle, and that's where the test stops. So we covered EVERYTHING that the test—that they're tested on. And drilled on it and drilled on it and drilled on it. So they were really set.

You see the large ups and downs [between ITBS and BST scores for an individual pupil] and you wonder, what is the value of testing? If you studied decimals and you tested decimals and had several tests taken over several days, and you took the average of those and threw out the real highs and the real lows and then maybe averaged the rest, then that would be a good test of what was going on. Especially for our kids, because our kids are not motivated to take tests.

It's an insane thing. I just don't know how educated, intelligent people can get into such a pickle like this and allow it to continue to grind on. Any reasonable person, they have to say, that's insane, that children go through that kind of testing. And then, for the district to say this is going to be important, we are going to be watching these tests carefully. Dr. Thorne has said that we hate to think of it, but they're going to be looking at the grades and looking at the teachers, and saying, "Who does the best?" He said, we hate to think of it, but that's what is going to happen. So this year, I walk on water. I'm a super teacher.

And next year? Who knows. I don't think...because I know the kids. I just know that they can just as easily fall down as do well. I mean they can't pass if they don't know, but they, it doesn't necessarily mean they will pass if they do know it.

Mrs. Samuel looked at the test scores as having few surprises.

Basically their scores reflect what they do, for most of them, in some ways. Cr, she doesn't care, and that's both on test and in daily work. She was in first grade two years, she went from a K.7 to a 1.1 to a 1.5, so in three years, she has not made a year's growth, on paper. Capability wise, she can, but this doesn't reflect what she's capable of. She just does anything she wants to. She doesn't care. She will be in third grade next year. I couldn't retain her even though she flunked spelling and language this year—she had 4's and 5's—because she had already been retained. And talking to mother is useless. She talks a good story and says she wants her to do well in school because she only got an eighth grade education and she wants her children to do better than that. And she tries to get Cr to care and to do better. But things that get sent home to be signed and returned don't get signed and returned. And I talked to mom and she says, "Oh yeah, I saw that." But she didn't make any effort to sign it and send it back. So, I don't know if Cr is ever going to do anything. Until someone lights a fire under her and she decides to care about it. I tried, all year, to get her to do something. And she could care less. It didn't bother her to sit against the wall. It didn't bother her to not make her day when everyone else in the class did. She just doesn't care.

R is above average student, so his scores reflect that. He's a real good student. Ra has an attitude problem. Her language could have been higher. [Referring to the tantrum thrown during the language test], that's Ra, and that's the language test, her thoughts are elsewhere. She certainly could have done better on that. She is one of the ones who has made [3.6 and 3.4 ITBS language scores] a lot of growth this year. She also was retained in first grade. She grew from 2.1 to 3.6, her scores look good, and I think they pretty well reflect what she is capable of, except for the language. She could have done better, but that day, she was not in the best of moods and it showed that. L also was retained in first grade. His scores don't show a whole lot of growth. He went from a 1.5 to a 1.9. [Because of his visual handicap, he] used the enlarged version of the test, though I don't really think he needed it. L will do any old thing regardless of what the directions are or whatever, and that is reflected here. He looked like he knew what he was doing when he filled in the circles. The aid spent a lot of time with him, particularly the first two days. She said that she was real impressed because he'd be right with us and he'd fill in a circle. And then she'd go back and look at them and what he filled in had nothing to do with whatever we were doing. That was on listening. His listening scores were the lowest of everybody in the class, K-5. And that's in line with him. He doesn't listen, he doesn't think.

Si, this was beyond her. Si has a lot of capabilities. And she's functioning pretty much on grade level. She has come up incredibly in reading. She couldn't read anything at the beginning of the year. She's more than half way through Reading Mastery II and perfectly capable of finishing it. She has trouble with the inferential type things. If it's not right there in the sentence she has trouble finding it. Part of that is the language [native language is Thai]. All of that reading on the test was beyond her. But she came up a year. She was K-7 last year to 1.7 this year. The vocabulary was the worst. The listening was the highest. She tried hard on the reading stories. I think she was one of the ones who didn't finish or at the end just guessed. She got confidently through the first couple of items and then it was beyond her.

K is capable of doing much better than she did. She is lazy. She doesn't want to have to read anything. A score of 1.9 is not too good. She doesn't attend,

> so I don't feel her scores reflect where she's at, grade-level-wise. She has mostly 1's and 2's in school. She got dropped to a lower reading group, because she acted like she couldn't do it. But she's more comfortable where she knows it real well and can skate. She doesn't want to put out the effort unless you really nail her on it, she won't. And this test, she filled in anything. I watched her on the reading and I had to really get on her about reading the paragraphs. She was the second or third one done, and there was no way, she was done before Ph [the best student in class] was [laughs]. There's no way. I asked her to go over them and she didn't do much. She played like she was reading them, but she really wasn't. So her score is probably a gift. She just guesses on most of it, and did well.
>
> But most of the kids' scores didn't surprise me. And in the school, on the whole, they were pretty pleased. Second grade did well. Ours looked good. They keep shooting for 2.9, which always irritates me, because [grade placement] is actually 2.8, because we took the thing in April. Dr. Thorne and Dr. Michael seemed to be very pleased, because our gains were all pretty good. Third grade was low, again. And it's always been that way. The last three or four years that I can remember.
>
> Not a whole lot was said about it. Well, one of the things that was said was because they did so well in second grade [the year before], to make a year's gain, they had to continue to do as well. It was brought up and both principals acknowledged that it was true. I've got these kids here at 3.5. In order for J to look good next year, he's got to have a 4.5. And that may not be possible. Third grade—there's a big jump between second and third grade. They recognize that. I don't think it helps our whole school average, but they recognize that fact. There wasn't...If there were comments made to individual teachers about what their scores looked like, I didn't hear about it. On the whole, they said they were pleased.

Several of the elementary schools performed poorly on the Basic Skills Test. The standard of 75 percent of objectives mastered by 75 percent of the pupils was met less (at about a 3:2 ratio) than it was attained, across grade levels and subjects. Some teachers explained that the district standard was almost impossible to meet without teaching test items. Four features of BST testing contributed to their explanation. The tests have poor content validity, even compared to the Scope and Sequence or required texts. "It's like they pulled questions out of the text by closing their eyes and pointing," one teacher said of the social studies test. Many questions are ambiguously worded, and if pupils miss more than one item wrong per objective, they fail that objective. The math test scoring routine misfired, contributing to erroneous scores. Finally, according to the teachers, after twenty hours or so of testing, pupils have little patience or energy left by the time the BST comes around in May. Most of the teachers and principals reacted casually to these results, believing the consequences that would ensue from them were minor.

Stage Nine: Aligning Instruction

Teachers spend summer with little thought to tests and scores. For administrators, however, test scores play a major role. This is the time when district administrators pour over reports of scores, reanalyzing them in many different ways

Table 3

HAMILTON ITBS RESULTS, 1986-1988

READING	GRADE 1	GRADE 2	GRADE 3	GRADE 4	GRADE 5	GRADE 6
1986	1.6	2.7	3.9	4.5	5.5	6.1
1987	1.5	2.9	3.4	4.9	5.9	6.6
1988	1.5	2.7	3.7	4.5	5.9	7.1
86-87 GAIN		1.3	0.7	1.0	1.4	1.1
87-88 GAIN		1.2	0.8	1.1	1.0	1.2
88 District	1.9	3.1	4.0	5.1	6.3	7.3

Average Grade Equivalent Scores

LANGUAGE	GRADE 1	GRADE 2	GRADE 3	GRADE 4	GRADE 5	GRADE 6
1986	1.8	3.1	4.0	4.7	5.4	6.1
1987	1.9	3.2	3.9	5.4	5.8	6.7
1988	2.0	3.1	4.2	5.1	5.8	7.0
86-87 GAIN		1.3	0.7	1.0	1.4	1.1
87-88 GAIN		1.2	1.0	1.2	0.4	1.2
88 DISTRICT	2.4	3.6	4.6	5.4	6.5	7.5

Table 3

HAMILTON ITBS RESULTS, 1986-1988

MATH	GRADE 1	GRADE 2	GRADE 3	GRADE 4	GRADE 5	GRADE 6
1986	1.5	2.9	3.7	4.7	5.5	5.9
1987	1.5	3.1	3.5	5.1	5.7	6.8
1988	1.8	2.8	3.9	4.8	5.9	7.4
86-87 GAIN		1.6	0.6	1.4	1.0	1.3
87-88 GAIN		1.1	0.9	1.2	0.9	1.5
88 DISTRICT	2.2	3.3	4.1	5.1	6.2	7.4

WORK STUDY	GRADE 1	GRADE 2	GRADE 3	GRADE 4	GRADE 5	GRADE 6
1986	1.5	2.9	3.7	4.7	5.5	5.9
1987	1.5	3.1	3.5	5.1	5.7	6.8
1988	1.7	2.7	3.7	4.7	5.7	7.0
86-87 GAINS		1.6	0.6	1.4	1.0	1.3
87-88 GAINS		1.1	0.6	1.2	0.5	1.3
88 DISTRICT	2.3	3.5	4.2	5.2	6.3	7.3

Table 3

JACKSON ITBS RESULTS, 1986-1988

READING	GRADE 1	GRADE 2	GRADE 3	GRADE 4	GRADE 5	GRADE 6
1986	1.6	2.9	3.8	4.6	5.8	6.4
1987	1.8	2.8	3.5	4.6	5.6	6.2
1988	1.6	2.7	3.7	5.1	5.6	6.9
86-87 GAIN		1.2	0.6	0.8	1.0	0.4
87-88 GAIN		0.9	0.9	1.6	1.0	1.3
88 DISTRICT	1.9	3.1	4.0	5.1	6.3	7.3

LANGUAGE	GRADE 1	GRADE 2	GRADE 3	GRADE 4	GRADE 5	GRADE 6
1986	2.0	3.2	4.1	4.8	5.8	7.0
1987	2.2	3.3	4.0	4.6	6.1	6.5
1988	2.1	3.0	4.1	5.1	5.9	7.0
86-87 GAIN		1.3	0.8	0.5	1.3	0.7
87-88 GAIN		0.8	0.8	1.1	1.3	0.9
88 DISTRICT	2.4	3.6	4.6	5.4	6.5	7.5

Table 3

JACKSON ITBS RESULTS, 1986-1988

MATH	GRADE 1	GRADE 2	GRADE 3	GRADE 4	GRADE 5	GRADE 6
1986	2.0	3.2	3.7	4.4	6.0	7.1
1987	2.3	3.1	3.6	4.5	5.5	6.1
1988	1.9	2.7	3.7	4.7	5.8	6.6
86-87 GAIN		1.1	0.4	0.8	1.1	0.1
87-88 GAIN		0.4	0.6	1.1	1.3	1.1
88 DISTRICT	2.2	3.3	4.1	5.1	6.2	7.4

WORK STUDY	GRADE 1	GRADE 2	GRADE 3	GRADE 4	GRADE 5	GRADE 6
1986	1.9	3.3	3.7	4.7	5.7	6.8
1987	2.2	3.3	3.9	4.7	5.7	6.2
1988	1.8	2.9	3.7	4.8	5.8	6.6
86-87 GAIN		1.4	0.6	1.0	1.0	0.5
87-88 GAIN		0.7	0.4	0.9	1.1	0.9
88 DISTRICT	2.3	3.5	4.2	5.2	6.3	7.3

and computing a variety of metrics and rankings they can compare with standards. Three measures dominate their attention: (a) arithmetic differences between grade placement and average grade equivalent scores on the ITBS, (b) the group gain attained on each subtest of the ITBS and the standard of one year's gain, and (c) the percent of objectives mastered on the Basic Skills Test compared to the standard of 75 percent.

District administrators then turn these reports over to the principals so that they can study them and plan programs to raise the scores and meet the district standards. Attention is focused particularly on principals whose school failed to attain one or more of these standards. They are told to find ways to increase scores. In addition, these scores are used in part to determine the principals' merit raises. By August, after the board receives the test score reports, the principals feel they are vulnerable even if scores are even slightly below standards or if only one of the many subtests is lower than others. A board member publicly chastises principals whose schools fail to attain the standards.

District curriculum coordinators also use the scores to plan district-wide modifications. Language arts is a telling case. A committee of coordinators and teachers meets over the summer to revise the district Scope and Sequence in language arts. They take the different forms and levels of the ITBS and determine which skills the test covers at the different grades. Then they revise the Scope and Sequence so that the skills the test covers at particular years are introduced the year before and mastered the year the test covers them. They align CUES and BST so that they conform to the ITBS and reinforce the same skills. They list the chapters and sections of texts and materials that teachers can use to teach the material and stress the skills.

During the summer the district completes its strategic planning project. The primary goal for the district is to "have 100 percent of students achieve at least one year of growth in reading, math, and language for each year in school." Although one might operationalize this goal in a variety of ways, the district rhetoric shows that they mean growth on the standardized achievement test scale, that is, average gain in grade equivalent scores equal to grade placement. In the strategic planning document is the strategy for attaining this goal: "We will develop and implement a comprehensive curriculum aligned with assessment measures consistent with our mission statement and curriculum."

When teachers return to school in August, the administrators reorient them to testing and test scores. In early meetings, teachers learn about the district's strategic plans and strategies and try to reconcile their own priorities. They hear presentations of test scores and messages that test scores are important and district standards need to be attained.

At Hamilton's initial staff meeting, a new assistant principal presents the previous year's test results. She states that the "impressive gains in reading test scores prove we're on the right track." On the CUES, Hamilton was above the district average in reading and matched it in language and math. There was no pattern of results on the BST, except that math seems to be the lowest subject area. The sixth grade teachers protest that the district testing office had mis-scored the math test. Ignoring their argument, she continues her report. The ITBS showed gains of one year in almost all grades and subtests. In most cases the group gains met or exceeded the average for the district as a whole. The fifth and sixth grade scores were "way above the national average." "We should feel good about the progress we've made. But we want to go all the way." There is some dismay expressed over the errors made at the primary level, and the principals promise that they will make an "error analysis" to determine where current instruction breaks down and where to make corrections in instruction and curriculum. They attribute the low primary scores to "where these kids are coming from."

"By the time they get to sixth grade, our students are competitive with any students in the country. This is not happenstance, this is real teaching." They urge the teachers to be tenacious, to improve even this good performance on the various tests. The capstone of the testing results is the Metropolitan, which they had administered to "the core of students," excluding those in transition classes. They cross-tabulated scores according to whether the pupils had been enrolled 120 days or more at Hamilton, but this breakdown revealed no differences between those above and below that criterion. Metropolitan scores were 1.7 for first grade, 2.9 for second, 4.2 for third, 5.2 for fourth, 5.9 for fifth, and 10.0 for sixth. Another analysis shows the percentage of pupils at each grade who were at the 50th percentile or above (ranging from 43 percent of first graders to 73 percent for sixth graders). The assistant principal reminds them that there is a "kid behind every one of those statistics," and "we need to aim for 100 percent at grade level or higher."

Later in the meeting, Dr. Thorne says that this year, "We are going to go for it," it being the designation of Hamilton as an A+ school. He describes one of the current award winners, whose principal reports that scores are very important to the judges. Like last year, the principals review the research literature on effective schools (criteria for being identified as effective relate to standardized achievement test scores) and positive school climate.

The third reorientation to testing that administrators give to teachers at Hamilton takes place in grade level teachers' meetings, where scores are reviewed and plans made to correct deficiencies. At the sixth grade meeting, the assistant principal rehashes the low math BST scores. Although teachers remind her that the scores are "bogus" (they had made up answer keys and scored the tests themselves before submitting them to the district testing office), she insists that teachers come back with a plan to fix math. At the fourth grade meeting, she addresses the problem of relatively low scores in language and proposes that they develop test-like practice exercises in mechanics and usage that they can use on a daily basis, the same way they already must do Systematic Review in math. Thus, for five or ten minutes a day, the pupils will work on exercises where they will review material on subjects, predicates, and other mechanics that have previously been covered in the curriculum. By such repetition, it is thought that pupils in Hamilton will improve the low scores in subsequent tests. Third grade teachers must patch holes in several subjects, working on the assumption that the low scores have an analog in either the programs or in the delivery of them.

These reorientations fall into a similar pattern. They focus teachers' attention on the tests and emphasize the importance of high scores to internal decision-making and evaluation by external audiences. By associating the teachers' efforts with the high scores achieved, the administrators attempt to persuade teachers that the tests are valid indicators of achievement.

At Jackson, test score reports occupy a minor part of opening meetings, although the teachers are quite aware of what the scores are. Mrs. Mitchell proclaims that they, too, will compete for A+ designation, although "Our scores will kill us unless we find some alternative means for proving the quality of our program." Jackson is on another mission quite apart from test scores. Not only are they refining their Whole Language curriculum, but they are launching a school-wide discipline program based on pupils' psychological needs rather than on reinforcing correct behaviors or extinguishing bad ones. Administrators and teachers become engrossed in discussing this, and the tests go by the way. Later, the teachers decide to use *Scoring High* again this year, because they believe that using this program was responsible for the the scores they obtained, which were higher than the scores from the year before. They are no more convinced than they ever were that the high quality instruction they provide shows up on the ITBS. They believe that, although current scores have staved off attacks on their special programs, they are still at risk for a district take-over if scores drop this year.

Summary

So ends the final stage in the natural history of the testing event and the initial stage in the next cycle. The meanings about external tests that teachers hold and the actions they take toward them change qualitatively through each stage. Tests play different roles at different points in the cycle. It was our intent in this chapter to show the texture of elementary school teachers' lives as they deal with the imposition of external tests. Comparing the testing activities at various stages to the beliefs of teachers about testing, it is possible to conclude that teachers commit substantial amounts of time and effort to the pursuit of higher scores even though those scores fail to represent educational attainment as they define it. They respond to pressures from administrators and external audiences, fears about loss of autonomy and feelings of self-efficacy, by engaging in activities that will boost scores, even as they acknowledge a sharp distinction between test scores and real achievement. By observing actions over a substantial period of time, one can assert that curricular narrowing occurs. Teaching to the test happens. Pressure on teachers to raise scores is a reality. Over time, because of external testing (although not by that cause alone), some forms of teaching come to resemble testing. Tests have peculiar content, form, and underlying assumptions about the nature of learning and curriculum that fit some programs better than others. Material that is not tested and forms of instruction that do not fit the test formats are abandoned unless teachers and principals take risks to preserve them.

Chapter Four: Assertions about Testing

External testing pervades the life of public schools and, despite the controversy it engenders, shows every sign of expanding its impact. From our 15 months of involvement with 2 elementary schools and a year of data analysis, we generated theoretical assertions in the following categories: local definitions of testing, the role of testing, and the effects of testing. Internal reliability checks, participant checks, multi-method confirmation, adequacy and sufficiency of evidence, and analysis of discrepant cases supported these assertions.

Definitions of Testing

The testing event we define as all the activities that make up planning for, administering, taking, and reacting to the scores from external tests.

Internal and external testing. Definitions in the literature differentiate testing programs into internal and external, roughly the distinction between testing programs the local school district initiates and those that some state agency requires. In this study, the emic distinction—the difference that makes a difference to participants—is finer. Internal tests are those initiated by teachers themselves and which are consistent with their particular conceptions of what and how they ought to be teaching. External testing is any assessment that the state, district, or principal mandates, whose administration means the teachers must interrupt their ordinary instruction to teach what the test covers and then assess the outcomes of that teaching. The ITBS is universally regarded as an external test. Most teachers regard the CUES and BST, which the state mandates but the district constructs so as to be consistent with district curriculum, as external testing programs. Although these latter testing programs are local, teachers view them as psychometrically flawed, unwarranted intrusions into the flow of their own curriculum, and biased toward those models of instruction that emanate from an overweening concern for standardized achievement testing. There are exceptions, however. Those few teachers who accept the district Scope and Sequence as the appropriate curriculum and teaching approach would define CUES and BST as internal tests. Even the Metropolitan Achievement Test, which the Hamilton principal mandated to evaluate the school's program, counted as external to many of the teachers there, for it failed to advance instruction or give teachers a basis for assessing pupils' progress. Like the teachers Dorr-Bremme and Herman (1986) studied, internal and external testing are "functionally independent" in this site.

To be internal, a testing program must incorporate teacher's conceptions of what ought to be measured and by what means. External testing programs wrest control over what to test—thus the question of educational values—and the decisions about methods of assessment away from those groups with the greatest awareness of local circumstances and the greatest interest in ensuring that the values are honored. On the other hand, external testing addresses, in theory at least, concerns of society for rational means to attain common standards by efficient methods. A disinterested third party, otherwise ignorant of actual transactions of schooling, ought to be able to interpret the results.

In the initial phases of the study, we used the term ordinary instruction to refer to curriculum teachers teach, the methods and materials they use, and the methods of assessing pupils' progress toward local goals, when these activities are relatively free from the influences of external testing. We modified the definition as we saw more activities the teacher accepted as ordinary instruction falling under the influence of external testing. For example, teachers perceived the Study Skills Handbook as a legitimate part of their curriculum, and many of them set aside a specific time to teach from it just as they set aside time to teach spelling. Yet the material in the Handbook, as one can readily observe by looking at its contents, offers activities to prepare pupils to take the BST and ITBS. Similarly, the teachers accept as part of ordinary instruction as they perceive it those math exercises in textbooks that mimic ITBS items.

We elected the inelegant but accurate label the packed curriculum to refer to the stupefying amount of material that the state and district expect teachers to cover in a school year. As the descriptions in Chapter Three revealed, a typical intermediate grade teacher's responsibilities encompass eight textbooks and three handbooks plus programs in computer literacy and drug resistance training. Each year the state or district adds something new but deletes nothing, for each addition attracts a constituency or even a niche in the bureaucracy. The school day is short, and time is a dwindling resource. Many things outside the teachers' control cut into the day. What the district refers to as "Specials"—physical education, music, and art—sop up 10 percent of the day. Children come and go for band, student council, and pull-out programs. Teachers' energies are not unlimited, either. Nor are they equally interested or competent to teach every part of the packed curriculum, a fact that eventually erodes their sense of efficacy and their notion that they themselves only must understand the "what" of teaching at a superficial level. The "packed curriculum" is related to Apple's (1982) intensification.

Stakes. By consensus (Madaus, 1987), the accepted definition in the literature of a high stakes testing program is one whose results trigger specific administrative action such as promotion or graduation, or one that pupils, teachers, or administrators perceive as likely to have consequences. A low stakes testing program is one without such perceived consequences, as when a state merely provides test data to districts so that they can diagnose or fix their own problems as they see fit. Our study highlights the perceptual and symbolic nature of the consequences of testing. To be high stakes, it is enough that a testing program has the power to shame, which publication in accessible media makes possible. The Arizona Department of Education reports ITBS scores for each school and grade level, and the press turns the scores into rankings. The motives to avoid the bottom position and the shame of occupying it are powerful. The distinctions of stakes in this study correspond to those of Wilson and Corbett (1989) when they discovered that publication of test score rankings in a public forum was just as powerful as the triggering of administrative actions. Some policy analysts accept the publication of rankings as effective ways of blowing the whistle on inadequate schools and shaming them into better performance or allowing parents to use these data as a basis for choosing among schools for their children (Rothman, 1989). Use by the federal government of wall chart indicators and proposed rankings of state results on the National Assessment of Educational Progress are the current manifestations of this philosophy. Its proponents often gloss over technical weaknesses of sampling, measurement, and analysis of the indicators.

Stakes of testing programs vary locally and from time to time. Several examples support this assertion. Conventional wisdom counts Arizona as a state with a promotional gates program. By state statute, schools must determine whether children pass from third to fourth grade and from eighth to ninth grade on the basis of district test results. This ought to be high stakes. However, the schools in our study barely paid lip service to this promotional gates policy. This finding supports Ellwein's (1987) analysis that the effect of promotional gates policy is largely symbolic. Merely having the policy in place makes the state or local district look tough. The teachers, however, finds ways of getting around them and using their own criteria for promotion and retention. It is also consistent with a symbolic interactionist conceptual framework: organizational structures, roles, and rules are only influential when participants take them into account.

Another example of the varying power of tests in Cactus District was the reaction to Basic Skills Test results in two different years. In 1988, principals felt that BST results lacked power and consequently paid the test less attention than they paid the ITBS. They correctly predicted that low scores on BST would have no consequences. In 1989, however, the central district and Board of Education suddenly took notice of BST results and began calling schools to account for failing to attain district mastery standards on them (even when those schools attained district standards on ITBS average grade equivalent scores and ITBS growth). It was no longer safe to ignore BST despite their acknowledged psychometric and content validity weaknesses. Principals and teachers felt that they could solve this problem by teaching items on the test itself, which they planned to do in the following year.

ITBS results trigger no actions in this district and contribute only a part to the principals' merit evaluations and nothing at all to teachers' merit evaluations. In fact, low scores often attract additional resources to a school. The organizational rules and social facts notwithstanding, the ITBS has tremendous power throughout the district and the state. Principals and teachers fear that the district will use low scores to dismiss or demote them, transfer them to other schools, or reduce their freedom to conduct special programs to which they are philosophically committed. These fears come about through rumors and subtle messages. They are difficult to verify. Central district administrators, as we showed in Chapter Two, regard these perceived consequences as unwarranted overreactions on the part of principals and teachers. In a grants economy such as public schools are part of, however, manipulation of symbols and status positions amounts to social reality. From past history in this district and present cases in other districts, principals surmise that administrators can be removed for reasons other than those the district makes public. Schools are taken over. Teachers are transferred among schools or grade levels against their wishes. Teachers and principals with low scores find themselves at a disadvantage. They actively seek alternative means of accounting for their educational outcomes and of establishing high status and acceptance in the community. They seek awards for themselves, their schools, their teachers, and solicit special programs, media attention, and any other form of recognition not based on test scores.

To honor participant meanings and distinctions, we must define stakes as the power test results have to trigger local administrative action, evoke feelings of shame, and decrease status within the organization, the latter being as "real" as the former. With that definition, we classify Cactus District as a high stakes testing environment with respect to the ITBS, ITBS growth, and, at times, the BST. There may be environments with higher stakes, such as districts with career ladder programs based on test scores, in districts where test scores trigger graduation or allocation of resources, or states that can take over districts with patterns of low test scores. To make such an assertion, however, one would have to understand participants' meanings in those environments.

Participants feel the effects of stakes, although they do not use that term, labeling it instead as pressure. One can understand participants' meanings for pressure in three different ways. First, pressure is phenomenological. It is a feeling people have and a drive to do what is necessary to avoid having their names or organizations associated publicly with low scores. The feelings might be competitive or ambitious, driving one to do whatever is necessary to have their schools attain high scores. Second, pressure is organizational, that is, codified in formal goals and evaluation systems of the district, manifested in this district's adoption of a goal that every child must gain a year's worth of achievement in a year's attendance. Third, pressure is transactional. In face-to-face meetings and written directives, persons at one level in the educational system encourage or stress the urgency and necessity of high scores to persons at the next lower level. The board pressures the district administrators, who also must consider the possibility that the public will turn down bond elections if scores are low. District administrators tell principals to ensure that their school averages are high or higher. These messages rarely urge schools to offer quality education; instead they focus more on raising the scores themselves. Principals pass the message to teachers who pass it along to pupils. Consistent with the conceptual framework of symbolic interaction, we find that persons at any level can reinterpret the message and break the chain if, for example, they deny the validity or importance of test scores. The chain is weak in any case, as pupil performance is hardly under teachers' complete control. The most teachers can do is teach well, cover the required curriculum, prepare pupils by whatever means to take the tests, and encourage them to do their best. From the actual scene of the battle pitting students against test items, each layer in the school hierarchy is further removed. It is this absence of control that contributes to feelings of alienation and pressure.

By official definition, norm-referenced tests are standardized achievement tests that compare one pupil's performance with that of similarly situated pupils nation-wide. Criterion-referenced tests compare a pupil's performance against a defined standard of competence. Although testing professionals often draw sharp distinctions between them, participants in this study treated them alike. Both types of test use closed-ended formats such as multiple-choice items and formal, standard rules for determining the meaning of the responses. Both can be internal or external; both can be used for individual or group assessment.

Some teachers preferred the criterion-referenced tests (CUES, BST and Study Skills Tests) because they were written by people close to home, not because of their referencing. Teachers complained about the length and difficulty inherent in norm-referenced tests, but criterion-referenced tests can also be long, difficult, and ambiguous. Teaching the test, teaching to the test, and outright cheating can contaminate the results of both forms of test. It is the power of the testing program and the use external audiences make of its results, not its form or referencing system that makes a difference to participants.

The concept of educational attainment that participants hold is broader than what achievement tests measure. When they define educational attainment, teachers name processes such as "helping kids develop an understanding of multiplication concepts" or "developing in students a love of learning."

Among participants, achievement itself encompasses outcomes of teaching, that is, a subset of educational attainment. Teachers fall into two types regarding their definition of achievement. Some name properties such as the pupils' "ability to do abstract kinds of problems," and others name properties such as "basic skills—you keep reviewing it, and it becomes part of their learning...what they've retained." Only the latter definition of achievement is consistent with models of learning implicit in standardized achievement testing. As Resnick and Resnick, (1989) recognized, those with theories of learning and teaching consistent with achievement testing models assume that knowledge and skill can be "decomposed" into independent, additive components (items on a subtest), the sum of which indicates the knowledge and skill as a whole. They believe that performance does not depend on the context in which it was expressed, and that "each component of a complex skill is fixed, and that it will take the same form no matter where it is used" (p. 11) or on which task it was originally based. They believe that successful learning is the matching of individual responses with those that someone else has previously defined as correct responses. They accept that technical considerations such as high reliability and low cost per unit of information ought to govern assessment of learning. They believe that disinterested and distant third parties can judge the adequacy of responses in standardized ways.

The tendency to think of learning in ways consistent or inconsistent with the models implicit in achievement tests was the only category in which teachers in our two schools differed from each other. Even in this respect, the correlation was not perfect. No matter in which of these categories they fall, all the teachers in this study believe that achievement tests reflect only a diminished and perhaps skewed portion of the set of all goals for which schools strive.

The most important set of distinctions that participants in this study identify regards the discrepancy between the indicator and the trait of achievement. Teachers are acutely aware of the weakness of standardized achievement tests (norm- and criterion-referenced) to represent adequately achievement as they define it. According to the beliefs of the teachers, the discrepancy is greater for certain kinds of pupils. For example, achievement tests may bore very bright pupils, creative or divergent thinkers, or pupils who "read too much into test items" and choose the wrong answer. For pupils with below-average intellectual ability, poor emotional stability, low self-confidence, weak motivation, and those who are having trouble at home or with friends or whose parents have neglected to instill in them the habit of perseverance, the test score fails to reflect their real attainment. Scores of young pupils are more apt than those of older pupils to fall short of their real achievement. Some pupils are simply better test-takers than others, and test scores reflect those characteristics rather than real achievement. Test scores depend on the pupils' intentions and effort, which vary along dimensions that have little to do with the qualities of teaching and learning they have experienced.

According to teachers' beliefs, achievement tests distort educational attainment because their content and format rarely reflect what has been taught in the classroom, particularly if the school's processes and goals diverge from the educational testing models. Teachers who aim for "authentic literacy" or conceptual understanding of math, or who base their teaching on cognitive psychology believe that their goals and activities match poorly the kinds of skill that achievement tests cover, that is, "rote-memory" or other "low-level skills." When the test covers what the textbook does not, the scores fail to represent locally defined achievement.

Features of the tests themselves (e.g., length, difficulty, ambiguity) also contribute to the discrepancy between the indicator and the trait of achievement. Teachers believe that children become tired, frustrated, and confused, and perform at less than their best or guess at the answers or fill in answer ovals at random. Teachers believe that the multiple-choice format limits the range of possible educational goals to those that can be easily tested, a problem that characterizes both norm-referenced and criterion-referenced tests. The restrictive and foreign environment of testing confounds pupils accustomed to working in groups or getting help from the teacher, a feature of testing that increases the discrepancy between the numerical value of the test score and the underlying trait of achievement. Teachers believe that the many tests the pupils must take exhaust them and lower their performance on tests taken later on.

There is a relationship between beliefs about testing and organizational role, position, and interest. Compared with teachers, district administrators gloss over the discrepancies between the trait and indicators of achievement. Teachers believe that achievement test scores ought to carry information about real achievement. District administrators acknowledge the inadequacies of tests but nevertheless study them "from every possible angle," looking for patterns of absolute or relative declines, differences among schools, grade levels, subtests, and differences between the performance and district standards. Central administrators overlook or choose to ignore technical flaws in tests such as unreliability of gain scores, ceilings on the amount of gain possible, the insignificance of differences between subtests, schools, and districts, and the unreliability of the tests (especially the district CUES and BST) themselves. In spite of these technical problems, central administrators encourage principals and teachers to raise scores that are low and promise the board and the public that schools will exceed earlier gains.

Unlike administrators, testing professionals, critics, and the public, teachers have unique access to "interpretive context," that is, the many other indicators (what was taught, what the pupil's state of mind was when he took the test, how hard he tried, how well he reads, computes, communicates, and performs on other tests and daily work) against which the meaning of the score itself can be judged. Other groups are more likely to assume a simple relationship between the trait and the indicator of achievement. However, we wonder if the public would accept this connection so readily if the actual items were to be made available. Test publishers, however, demand security of items. When the local newspaper revealed an item from the ITBS about the relative sizes of fruit varieties, public sentiment grew that achievement test items can be something less than logically related to commonplace notions of educational attainment. The public lacks awareness of the technical features of achievement testing, information such as the degree of measurement error, which also contributes to a kind of public mystification about testing. The closer one is to the actual scene in which learning takes place, the less likely one is to believe that achievement tests yield adequate information about real achievement.

Achievement growth. When teachers and most others think about pupils' academic progress, they typically use the discourse of gains or growth from year to year in achievement. Perhaps they imagine that children ought to know more things, have more skills, be able to reason better or read more sophisticated books in fourth grade than they did in third grade. Participants in this study think no differently when they imagine achievement growth in the abstract. To measure this abstract conception of growth and hold schools accountable for it, district administrators turn to the ITBS. As official policy, achievement growth is the difference in a school's average ITBS grade equivalent scores from one grade to another across years. That is, the average grade equivalent score among third graders in 1988 subtracted from the average grade equivalent score of fourth graders in 1989. Cactus District is not far off the definition of achievement growth that the Arizona Department of Education suggests, which uses differences in percentile ranks on the ITBS from one grade to another. The district holds a standard for its schools' production of achievement growth: a year's difference in ITBS grade equivalent scores, averaged across pupils in a grade level. In the District's strategic goals, this definition becomes the official policy: each child should gain at least a month in achievement for every month in attendance. The district holds some schools to higher standards. Administrators believe that schools should outperform in yearly growth the growth they made the previous year. Thus a school whose second graders "grew" an average of 13 months from first to second grade should "grow" the next year by at least fourteen months between second and third grade, all in ITBS grade equivalent terms. If they grow only one year between second and third, the school has failed to meet the district's standard of excellence. A school's growth statistics influence its principal's merit evaluation and pay raise.

The district's growth standard has several functions. It has symbolic value, advertising to its board and patrons that its schools are pursuing excellence and upholding high academic standards. It diverts public scrutiny from low grade equivalent scores to adequate growth scores. It provides a means for traditionally low-performing schools like Hamilton and Jackson to look accountable and excellent despite their concentration of disadvantaged pupils and the well-known relationship between disadvantaged populations and low test scores. Hamilton's principal pointed out with pride that, although they started lower, their pupils grew more on reading comprehension than any other school. By attributing the growth to their good teaching, he was able to boost the morale of the teachers and spur on their efforts on the next set of tests. Jackson's principal was able to ward off criticism by pointing to her school's growth statistics. By using the growth statistic, school staff can, at least temporarily, make it look as if their efforts are paying off, even when local average grade equivalent scores fall below national averages.

Psychometricians have long debated the merits of grade equivalent scores, detractors noting that they are based on different distributions (and different variances) at different grade levels. Therefore apparent gains may be spurious. It is generally acknowledged that tests at particular grade levels have ceiling effects, so that the possible amount of growth is limited. Schools already high in the distribution of national school scores bump into the ceiling. For teachers at these schools, the district standards for exceeding previous year's growth seem (and are in fact) impossible to meet. Over time, many teachers are becoming aware that the growth metric is a zero-sum game, in that second grade teachers who may profit from it one year severely jeopardize third grade teachers the next.

Other psychometric concerns come into play in the local definition of achievement growth, although they are yet to emerge in the consciousness of participants in the study. For example, even simple gains computed on individual pupils from one testing time to another have low reliability. Different groups of pupils take the test from one year to the next so that the averages between the two years can not be attributed to underlying real differences in the quality of instruction and learning (Berk, 1988; Cook & Campbell, 1979; Cronbach & Furby, 1970).

If any of these psychometric and statistical issues is true, then the official interpretation of achievement growth lacks meaning. Nevertheless, the district holds school performance up to this standard and assails those principals who fail to attain it.

Test utility. Like teachers in most surveys about testing practice, teachers in this study define external testing programs as useless in advancing instruction or assessing pupil progress or program success. Some researchers have interpreted teachers' failure to use test results to their lack of knowledge about testing. This explanation fails to account for the data at hand. Formal, propositional knowledge was in short supply among not only teachers, but other groups as well. However, it is the failure of existing achievement tests to represent their definitions of achievement and attainment that best explains teachers' beliefs about test utility and their lack of use of testing results.

Although teachers believe that external test results are of little use in advancing instruction or assessing pupil progress, they believe that other groups in fact use the scores. Teachers believe that decision-makers at the district and state levels use their test results against the interests of teachers and pupils; that is, to divert resources to meet problems that tests purport to identify, to direct attention to the skills that tests cover, and to reduce the span of autonomous actions teachers have.

Administrators use the district testing program as organizational tools to reward, punish, cajole, and control, regardless of the information scores carry about real achievement. They see the tests as ways to make sure all schools adhere to the District Scope and Sequence. As Cronbach (1984, pp. 347-348) wrote, standardized testing is one management strategy that helps administrators "to press persons at lower levels to strive harder" and "influence what will receive emphasis." Administrators use scores to manage the impression of the district that external audiences have (Edelman, 1976). High test scores can counter the demands of special interest groups or answer any charge that the schools are deficient. As a superintendent in a nearby district claimed, "I like the ITBS because we score high and that keeps the public and the board off our backs."

Test preparation. At the sites we studied, administrators make materials available, remind teachers of the importance of high scores, and then seem to wash their hands. Then, the teachers decide what kind of preparation to use and how much to do. In other sites, the district or school administrators set guidelines or specify rules; for example, either specifically forbidding or requiring use of *Scoring High.* Among elementary school teachers throughout Arizona who responded to a questionnaire (Nolen et al., 1989), 80 percent reported being encouraged to raise ITBS scores. Thirty-two percent reported that someone required them to prepare pupils to take the ITBS. Twenty-eight percent reported starting their test preparation activities two months or more before the ITBS.

The meanings of test preparation and the actions teachers take to prepare pupils to take external tests show interesting variety. What follows is a typology of meanings in action.

Some teachers do nothing. Perhaps they elect to do no special preparation because they are especially committed to some form of ordinary instruction, or because they believe that tests fail to measure defensible definitions of achievement, and because they have no fear of consequences of low scores. Following the categories of Mehrens and Kaminski (1988) and Shepard (1989), scores pupils attained in classes such as this would be valid, as standardization samples did no preparation beyond simple test-taking tips and taking the practice tests. Of the Arizona elementary teachers in the Nolen et al. (1989) survey, 12 percent reported doing no preparation for the ITBS. Most teachers admitted encouraging pupils to get a good night's sleep and breakfast before the test and to try hard on the test itself. In this site, encouragement did not extend to conducting pep rallies, as it does in some settings.

Teachers train pupils in test-wiseness. This method of test preparation seems to be acceptable to participants and testing experts. Mehrens and Kaminski (1988), for example, list such preparation among legitimate practices. Shepard (1989) claims that because norming samples also had some training in how to take tests before they took the ITBS, that similar training in ordinary test-takers will yield valid results. In the Nolen survey (Nolen et al., 1989), 60 percent of elementary school teachers reported teaching test-taking skills.

What the experts think of when they think of such training is probably just a mere shadow of what exists in Cactus District, however. In this study, district personnel have built systematic test-taking techniques into units of curriculum in what Hamilton's principal called "survival skills." The Study Skills Handbook in Cactus District served this purpose. Thus, does test-taking skills become part of the taken-for-granted curriculum and instruction come to resemble tests.

As Haladyna et al. (1989) pointed out, when some schools train pupils in test-taking techniques and others do not, comparisons between their sets of scores can not be explained as relative quality of teaching and learning. Thus, the legitimacy of such preparation to yield true pictures of a school's program or to serve as a basis for policy decisions is problematic.

Adding fuel to the controversy, various syntheses of research on the effects of such programs (Samson, 1985; Scruggs, White, & Bennion, 1986) reveal estimates of effects ranging from one-tenth to one half standard deviation. Although these effects are small, even the smallest can look like a month's advantage in grade equivalent scores, which might be enough for our participating teachers to demonstrate to external audiences and critics that they are doing their jobs. Or, from the external perspective, the effects are large enough to confound valid comparative interpretations and policy decisions.

Teachers in our study prepared their pupils for tests by reviewing the content of ordinary instruction, sequencing topics so that those the test covers would be taught prior to the test, and teaching content that they know the test covers. Experts define such activities as teaching the content domain of the test or teaching the objectives of the test (Shepard, 1989). Two examples at Hamilton illustrate this category. Mr. Armstrong repeatedly drilled material from ordinary instruction, in this case operations with decimals, that the test also covers and reversed the order of geometry skills and metrics, because the test covers the former and not the latter. Mrs. Samuels taught new material on contractions and compound words that the test covers but her curricular program does not. The Arizona survey (Nolen et al., 1989) revealed that 66 percent of elementary school teachers teach or review topics the ITBS covers.

There are at least two alternative reactions to this category of test preparation. Many define this activity as simply good teaching, systematically drilling and repetitiously reviewing sets of skills and objectives that almost everyone can agree are basic skills. One school in a different district in the metropolitan area has adopted for its entire curriculum sets of worksheets that prepare children for the ITBS. It calls itself a traditional school. Some educators argue that these skills on the ITBS and on worksheets that teachers use to prepare children for the ITBS adequately represent the entire construct or trait of achievement.

Mehrens and Kaminski (1988) claim that the items constitute only a sample of items from a universe of content and a broader construct of achievement. The score on the set of items must support an inference to the broader construct. By teaching items on the test or practicing test-like worksheets, the inference from the ITBS scores to the construct of achievement is no longer valid. The inference is "polluted," (Haladyna et al., 1989), and the relationship between indicator and construct is altered, tainted, or distorted.

Shepard explained it this way (1989, pp. 12-13):

> The original standardization sample did not have the benefit of such focused instruction. Students in the norming sample were apparently learning the tested content and other things as well when they took the unannounced test...[Inferring to a general construct of achievement from a particular test A] answers the question, "How would students who performed at percentile X on test A, do on test B?" As soon as schools begin to tailor instruction to a particular test [or teach to the specific objectives of the test], these equivalences no longer hold. As far as the public meaning of test scores is concerned, however, there is an implicit assumption made that these equivalences are true...But once curriculum has been aligned to the local test, there is no guarantee that apparent gains generalize to the non-taught-to tests...A local district that used a test but maintained a broad curricular focus beyond the test domain would be at a disadvantage in such comparisons [with schools that engaged in test-preparation activities].

Even changing the sequence of topics with an eye on what the test covers supposedly tips the scale toward focusing instruction on the specific objectives the test covers. If the review is effective, one has changed the relationship between the indicator (ITBS) and the broader construct of achievement.

Teachers coach pupils in specific formats of the ITBS. Teachers we studied used *Scoring High on the ITBS* and materials the district writes and distributes, that mimic the format and cover the same curricular territory as the ITBS. Some writers refer to this activity as "teaching to the test." Based on their content analysis of *Scoring High on the ITBS* and the items on the ITBS themselves, Mehrens and Kaminski (1988) concluded that using these materials had the same effect on scores as would administering a parallel form of the ITBS and explaining all the answer options to the pupils. In the survey of Arizona educators (Nolen et al., 1989), 41 percent of elementary school teachers reported using commercial test-preparation programs.

Several analogies have been used to explain the resulting effect. Mehrens and Kaminski (1988) compared the increase in test scores one obtains by using test preparation activities with practicing the Snellen eye chart and attributing the score after practice to real improvement in one's eyesight (and need for glasses). Baker (1989) referred to the rubber ruler, that is, that the test measures the construct one way in the absence of test-preparation activities and another way in their presence. One might also look at the increases in ITBS scores after test preparation as an inflation of a nation's currency without an underlying change in real wealth.

For testing experts like Mehrens and Haladyna, the preservation of the integrity of the inference from test score to the construct of achievement is the foremost concern. They refer to test-preparation activities that distort or "pollute" the inference as unethical and illegitimate. Haladyna went so far as to recommend that the Arizona Department of Education ban districts' use of *Scoring High*. By seeking to be the "critical reality definers" (Ball, 1987), testing experts stress issues of test validity and ignore the political issues lying behind educators' decision to teach to the test.

Teachers view these activities differently from experts and external audiences. For example, we observed that some teachers prepare pupils to take the test by boosting their confidence and inoculating them against the emotional effects of taking the test. Other teachers engage in test preparation as a means of enhancing their status or avoiding embarrassment.

Central district administrators regard test preparation as an irrelevant overreaction on the part of teachers and principals to unwarranted fears of the imagined consequences of low test scores. Principals and teachers regard test preparation as a means of self-defence. If scores are increased by means of these activities, the time will be justified by the end result of preserving autonomy over programs to which they are philosophically committed. They view these activities with cynicism. They hold their noses when doing them. They do not look at them as changing the underlying quality of instruction or learning. They see them as ways to boost the indicator without having much effect on the trait of achievement. Test preparation is the only control teachers believe they can exert over test scores, as they are so heavily influenced by pupils' socioeconomic status, ethnicity, and first language, and because tests are so distinct from ordinary instruction. According to Pechman (1985), test preparation is the only way people have of reconciling conflicting demands on their time and energy.

At Jackson, test preparation is a form of resistance. The staff is committed to Whole Language forms of instruction, and they perceive achievement testing as antithetical to their philosophies. They believe that the content of tests and the forms of items render the scores meaningless as indicators for the form of education they espouse. Yet they also believe that district and other external audiences accept the achievement test score as an adequate representation of school quality, and given a sufficiently long string of low scores, will act against the school to reduce its autonomy. Teachers use *Scoring High on the ITBS* because they believe it is the most effective and efficient method of test preparation. Using its materials, or the nearly identical materials that the district distributes, will produce the highest scores at the least cost of time they have to spend in ordinary instruction.

To chastise teachers for unethical behavior or for "polluting" the inference from the achievement test to the underlying construct of achievement is to miss the critical point. The teachers already view the indicator as polluted. They see its inadequacies in terms of content validity, reliability, the influence on scores of socioeconomic status and ethnic group. They see how district administrators and the board use its results fallaciously, trying to make curricular changes based on small and unreliable differences in test scores, pressuring principals when last year's growth is not as high as they would like. Why should they act against their own interests and those of their pupils as they define them? Why should they care about preserving the integrity of the inference from indicator to construct when they believe it already lacks integrity? When the pressure is on, teachers will look for ways to "pollute" the indicator.

The fallibility of indicators used as measures of accountability is well-known. Campbell (1979, p. 85) hypothesized that "The more any quantitative social indicator is used for decision-making, the more subject it will be to corruption pressures and the more apt it will be to distort the social processes it is intended to monitor." Ginsberg (1984) cited numerous examples supporting this hypothesis. She showed that tests, checklists, and other measurement devices whose developers originally intended that they measure such things as achievement, psychological well-being, unemployment, and crime were all compromised when some governmental body chose to use them as indicators of accountability of service agencies such as schools, mental health centers, social welfare departments, or probation offices. When the governmental body, for example, puts pressure on service providers, they and their administrators find creative ways to boost the numerical value of the indicators without changing the underlying quality of the services. The implications for school accountability are these: Standardized achievement tests are designed as measures of individual pupil progress in relation to national peers. If external audiences use ITBS scores as measures of school effectiveness and accountability and as triggers for reform, school personnel will focus their efforts on improving the scores without respect to, and to the neglect of, other equally plausible and valuable outcomes. The boosted indicator will not likely generalize to alternative indicators, such as the number and quality of books the children read, their writing, projects they undertake, or even to other achievement tests. When an indicator is so fallible that it changes in relation to short-term test preparation and test-wiseness training, or the social and ethnic composition of the population, it is worth little in public debate over school effectiveness or in the disbursing of rewards and punishments from society.

There is little difference of opinion on the legitimacy of "teaching the test," that is, providing practice on actual items of the ITBS or a parallel form of ITBS, a practice that can increase scores by as much as six months or more. Although teachers in this site did not teach items and test security was fairly stringent, opportunities to do so were present. Teachers felt it was improper and unprofessional to do so, although increasing the stakes might make them think again (Glass, 1989). They did, however, take long looks at the contents of subtests they were to administer later in the week and organize systematic review of material they found the later tests covered.

Cannell (1987) and others define this activity as cheating, for the fallacious scores that it produces rob the public of accurate information about schools. He argued that apparent gains in achievement scores over time were not due to any underlying improvements in teaching and learning but to inflated test norms, districts' practice of selecting testing programs that made them look good, teaching to the test and outright cheating by teachers and principals. In a study that Mehrens and Kaminski (1988) reported, 11 percent of the teachers surveyed reported cheating on standardized tests. Approximately 10 percent of Arizona elementary school teachers studied by Nolen et al. (1989) reported that they taught items from the current or previous year's ITBS. Twenty-six percent reported teaching vocabulary words that would appear on the test, both practices that Cannell would call cheating and Nolen et al. (1989) called "obviously unethical."

To prepare pupils for the Basic Skills Test, some teachers taught the test itself. No one specifically forbade such a practice, and handing out the tests early in the year almost encouraged it. Teachers in our study said that teaching the items on the BST was even necessary—that it was the only way to pass the test at mastery levels. Their reasons had to do with the poor content validity and the psychometric deficiencies of those tests. The tests are poor maps (McLean, 1989) of the contents of ordinary instruction, even as defined by textbooks. "It was like they pulled out sentences at random from the text," so one teacher described the BST in social studies. Advocates of criterion-referenced tests might justify practicing items as tantamount to teaching, but the public views a statement that sixth graders attained mastery on the social studies test as equivalent to good social studies teaching and learning. Therefore, a universe of content beyond the collection of test items is implied, even assuming that what good social studies instruction strives to attain can be thought of as a collection of "content."

Role of Testing

The role of testing changes over time in relation to the proximity of external tests and the time of year. A natural history of the testing event serves to organize participant actions and meanings with respect to testing. Actions and meanings of teachers and others change through the year in recognizable stages before, during, and after the test and the publication of test results.

At Stage One, teachers and principals confront the "packed curriculum" they are expected to cover during the year and must reconcile with their own goals. They must recognize that the demands exceed time and energy available. The role of testing at this stage is to suggest a priority to teachers about what they can safely omit or neglect in favor of content they already know the tests will cover. Recitations of last year's test scores and reminders of what happened as a result of the scores set in motion a series of actions by staff to avoid those consequences and public failure the next time. District administrators communicate the message that test scores are important and teachers and principals should make sure that they are high.

At Stage Two test results structure schools. Along with other sources of evidence and judgment, test scores determine the possible learning opportunities or the face-to-face groupings of pupils. These help define for the pupil what he is and what he can possibly become.

At Stage Three external testing recedes in educators' attention in favor of ordinary instruction, and tests that teachers use function to advance instruction and monitor pupil progress toward goals teachers accept as legitimate. When they view CUES as unwarranted intrusions and departures from ordinary instruction, teachers "perform" the assessments in ritualistic ways. Confronting the packed curriculum, teachers are apt to neglect those parts of it that do not appear on external tests. Teachers who neglect curriculum that is on the external tests or Scope and Sequence in favor of content or modes of instruction that they think are more educationally sound do so at some personal risk. To do the required curriculum and their personal one requires enormous energy and more time than is available. Tests also play hidden roles. When modes of instruction mimic modes of testing, teachers may not even recognize that they are teaching to the test merely by teaching. When formative tests such as CUES mimic summative, external tests such as ITBS, merely taking the formative test is equivalent to practicing the summative one.

At Stage Four, administrators pass along messages to teachers about the efforts they should make. These messages are not about improving education but about attaining high scores on the upcoming tests. As a result, teachers begin to orient ordinary instruction to the contents and formats of upcoming tests and plan for actions that they will take to prepare pupils for the tests. They make choices that align instruction to the tests. For example, teachers drop writing programs in favor of instruction in writing mechanics, drop instruction by math manipulatives for worksheets.

During Stage Five, from one to four weeks before the ITBS, teachers reduce substantially the time and energy they normally spend in ordinary instruction so that they can prepare their pupils for the test. They do this by reviewing what they normally cover, altering the sequence of topics, explaining or teaching new content they know the test covers, coaching pupils in test-taking skills and specialized formats, and attempting to build a sense of competence and self-confidence. Test preparation for tests other than the ITBS has similar qualities but differs in degree, in keeping with the relative power of the tests to cause shame or trigger district actions. Ordinary instruction diminishes, and time spent on untested material (e.g., writing, science, social studies, computer literacy) nearly disappears.

Testing at Stage Six consists of taking the standardized achievement test, resting from the harsh demands it places on students, and preparing for the next test in the sequence. Little ordinary instruction goes on.

During Stage Seven, something about the grind of preparing for and taking the test seems to necessitate a recovery phase. Freed from the demands of testing, the teachers use this time to restore their own priorities. But less energy is available to pursue them vigorously. Between the time they take the tests and the time the testing company reports the results, schools must reorganize for the subsequent year. Here the schools use the same mechanisms they used at the beginning of the year, but only year-old test results are available to help them.

In Stage Eight, scores come back from the testing company and teachers attempt to process their meanings and reconcile them with the other indicators they have of pupils' achievement and attainment. Principals read the test score reports as only the first of many they will receive and make predictions about the district's reactions.

In the final stage of the cycle, Stage Nine, district administrators reanalyze scores in many different ways and compare them with district standards. Not all of these patterns make sense, for example, when they try to interpret statistically trivial differences between schools as reflecting real differences in school quality, or insist that programs be changed when low scores were really due to the inadequacies of the tests themselves. They use results in combination with other indicators in determining principals' merit evaluation and salary increase. District administrators then turn these reports over to the principals so they can study them and plan interventions to raise the scores and meet the district standards. They pay particular attention to principals whose schools failed to attain one or more of the standards. They encourage principals to find ways to increase scores.

District curriculum coordinators also use the scores to plan district-wide modifications. The new goals and curriculum modifications make ordinary instruction more consistent with external testing and leave less time for teachers or principals to choose what to teach and how to teach it. Those who ignore pressure from the district to narrow offerings and align them more closely with the test put themselves at risk (either in reality or in their own perceptions).

When teachers return to school in August, the administrators reorient them to testing and test scores. Teachers learn about the district's plans and strategies and try to reconcile their own priorities with the required curriculum, which is even more packed than the year before. They hear presentations of test scores and messages that they need to raise them even higher and meet district standards. These reorientations focus teachers' attention on the tests and emphasize the importance of high scores to internal decision-making and evaluation by external audiences. By associating the teachers' efforts with the high scores achieved, the administrators attempt to persuade teachers that the tests are valid indicators of achievement.

Impacts of Testing

Our causal interpretations rest on the patterns of changes and co-occurrence we observed over time, on plausible attributions of participants in the study, and on logical accounts (plausible given the evidence) of the differences between what we saw in a high stakes testing environment and a range of alternative possibilities.

External testing reduces the time available for ordinary instruction. The most glaring impact of external testing in high stakes environment is on instructional time. In the packed curriculum, time is a non-renewable resource and is systematically reduced when external testing is introduced and stakes raised. We attempted to estimate the time the schools spent in preparing for, taking, and recovering from external tests by examining time allocations in the classrooms we observed and coalescing statements of teachers. In Hamilton's sixth grade classes, for example, teachers have approximately 30 hours per week of teaching time exclusive of specials and breaks, but not exclusive of pull-outs and miscellaneous programs. Time requirements of external tests themselves occupy about 18 hours. Based on the classes we observed, on the average, the teachers spent three hours of test preparation for every hour of external test administration. Recovery from testing, time when no ordinary instruction took place, occupied approximately two hours for every hour spent in external testing. The sum of these separate estimates exceeds 100 hours, or somewhere between 3 and 4 weeks of school time. Time teachers spend in internal testing in the course of ordinary instruction is additional. In the primary grades, the number of hours of external testing is 13 hours, with about the same ratios of testing time to preparing and resting time. The amount of time available to primary grade teachers is also less, about five hours per day. These estimates only encompass time teachers divert from ordinary instruction and not the structural and hidden effects we consider below.

Participants attribute the amount of time they spend in test preparation to the power of the test to evoke consequences. Teachers allow pupils to rest from the tests because they believe the tests hurt pupils. We find these explanations credible and consistent with our observations. Give teachers the hypothetical question whether they would use time this way in the absence of high stakes tests and without question they say the they would find other ways to use the time.

The test burden is likely to increase. The Arizona Department of Education proposes to add tests and assessments of "essential skills" that will be administered three times a year. All the other tests have constituencies that make it unlikely any will be deleted.

Testing affects what elementary schools teach. In high stakes environments, schools neglect material that the external tests do not include. Except for individual teachers with deep commitments to science, writing, or social studies, there is a strong tendency among teachers to spend most of their available time (including what little discretionary time they have) on reading, word recognition, recognition of errors in spelling, usage, punctuation, and arithmetic operations. Reading real books, writing in authentic contexts, solving higher order problems, creative and divergent thinking projects, longer-term integrative unit projects, computer education and such are gradually squeezed out of ordinary instruction—a joint effect of limited time, packed curriculum, and the imposition of external testing.

With the exception of a few teachers, science at the intermediate grades looks more like reading all the time. Teachers feel they cannot afford to take the time required to set up science activities or do divergent problem-solving. Hence, they spend the time having the pupils read the text and answer the questions at the back of the chapter and take the unit tests. As the external test approaches, time regularly allocated for science is siphoned off for test preparation. Similar things happen to social studies. Time some teachers formerly spent on writing projects, they later devote to instruction in formal grammar under the threat of the test. In a neighboring district, teachers say they have given up teaching science and social studies since they have designated themselves a "traditional school" and focus exclusively on the basics, that is, what the ITBS covers.

The decline in teaching of science and other "nonbasics" or untested subject matter can be explained this way. Regardless of the stakes of external testing programs, there will always be some teachers who teach science and teach it well. There always will be teachers who neglect it entirely or teach it poorly, regardless of stakes. But over time, as stakes increase, the trend will favor those who sacrifice science to spend available time on tested skills.

What is the effect of this hashing of science, social studies, and writing? Cognitive psychologists note that learning of details rests on prior cognitive schemata or prior learning. High school students who have had no prior knowledge are less likely to learn new material efficiently or effectively. Over time, the public will grow dissatisfied with lack of understanding of government, geography, economics or science-related issues that high school graduates exhibit and these graduates' failure to assume technological careers. This is already beginning to occur. Slighting content for skill assumes the two are separable, as if one can think without thinking about something. Science and social studies give students something to read, write, think, and discuss, and provide avenues to provoke or build on pupils' interests.

Many have argued that the greater time schools spend on teaching basic skills because of external testing is worth the loss of science, social studies, and writing. However, as our observations revealed, focusing instruction on tested material also slights (in topic, complexity, and form) reading, math, and language beyond those parts of reading, math, and language the tests cover. To illustrate, by intensively reviewing geometry skills, Mr. Armstrong neglected metrics and pre-algebra.

Whatever one decides are the merits of decreasing one subject in favor of another, one can hardly support the merits of curricular decision-making that occurs implicitly, without serious reflection or discussion. Curricular alignment with test contents preempts debate among interested parties and reasoned and moral decision-making about what schools should teach. The narrowing of curriculum observed in this study fits the interpretation of Corbett & Wilson (1989) and Darling-Hammond & Wise (1985) that external testing results in the substitution of means for ends.

External testing encourages use of instructional methods that resemble testing. One looks at a worksheet and an item from a standardized test of achievement and finds them nearly indistinguishable. Both call for the pupil to select among alternative options the one an outside expert has decided in advance is correct. Over time and with increased stakes, teaching becomes more test-like. Consider the consequences of scores on the ITBS language test on Hamilton's instruction. Hamilton's principal added Systematic Review of grammar (exercises that require pupils to identify or supply the correct answer) rather than increased opportunities for pupils to write, better preparation of teachers to teach language or writing, aids to help grade papers, or a different text or set of teaching materials. As a result of Systematic Review, there would be less time for the teachers to pursue alternative teaching forms. Consider Jackson's decision to drop the language CUES pilot, even though it was more consistent with the school's philosophy, because it failed to correspond to the forms of instruction and assessment in the BST and ITBS. Consider how teachers dropped Math Their Way as the external test neared in favor of speed drills and worksheets similar to the tests themselves.

Both tests and test-like teaching methods presume a relationship between subparts and the whole (like reading skills and reading). Subparts of reading are skills such as identifying word endings, identifying short and long vowel sounds, identifying main ideas in short passages matching syntax of test questions and reading passages. Some curricula emphasize these skills, and so do standardized reading achievement tests. Holistic programs emphasize arranging authentic interactions between pupils and texts; their proponents believe reading "skills" emerge from rather than provide a necessary base for real reading. Reading instruction based on principles of cognitive psychology also rejects the behavioristic building block model (that basic, lower order skills must be in place before the child can proceed to higher order problem-solving, comprehension of texts, and application). Cognitive-constructivists models (Glaser, 1984; Peterson, 1989) emphasize instruction that relates new knowledge in a meaningful way to the knowledge pupils already have, on the assumption that human beings construct knowledge out of their own experience. Models that base instruction on pupil interests or that emphasize enrichment are also poorly represented by achievement tests.

Tests are not value neutral or equally fair to all programs. The higher the stakes, the more instructional methods and materials will be test-like. Jackson serves as a limiting case in this assertion, as teachers and the principal have alternative commitments and are willing to risk the consequences of external testing programs.

Through the CUES and BST, the district promulgates a kind of mastery or minimum competency teaching model. They require repeated review of minima, such as using the clock and coins in arithmetic problems, reading numbers from graphs and charts, changing word problems into arithmetic algorithms, and correctly placing commas in personal and business letters. Repeating this instruction and testing it in grade after grade has several effects. So do minima become maxima. By stressing perfect mastery, this approach ignores the fact that pupil's boredom or carelessness or poorly worded questions or poorly drawn charts and pictures can often be a better explanation of imperfect performance than any underlying lack of comprehension of skill on the part of the pupils. Third, there is a great tendency for the teachers to stop instruction where the Scope and Sequence and the competency test stop, whereas, as one teacher said, "You could go on forever, because there is really no limit once the children get going on something they're interested in." There is less time and teacher energy to pursue divergent learning or enrichment activities.

External testing affects school organizations by placing general boundaries on placements and instructional opportunities. In the schools we studied, scores indirectly and directly influenced decisions to place students in homogeneous groupings. Teachers and administrators used achievement test scores along with other indicators, class work, and teacher judgment, to place children into transition classes, for example. At Hamilton, removing a child from a regular third grade class into a transition third grade slowed the child down to a pace or backed him up in the curriculum to a place where he could be successful. Such a decision also removed him from an environment where he could learn from more able children or which might press him to higher levels of effort. The move was permanent for his elementary career, for no children were accelerated or otherwise made up the ground once lost. Another effect of such a move was to remove from the regular third grade teacher's average one of the lower scoring pupils.

At both schools test scores were used directly, and with little room for interpretation, to place children in programs for the gifted and into a highly stratified junior high school curriculum. Despite (by teachers' definitions) the weaknesses of the test scores to represent adequately the underlying trait of achievement, particularly for disadvantaged and minority students, the test scores carried most of the weight in the decision. As a result, the children denied special services for the gifted and those entering a low homogeneous stratum in junior high lost appropriate educational opportunities because of test scores. In decisions about special education placement, teachers at both schools accorded more status than they ought (Smith, 1982) to scores of the psychologist's and special education diagnostician's test batteries. The numbers the specialists produced had a kind of magic quality for the teachers and resulted in decisions that further structured pupil careers.

External testing has hidden structural effects on ordinary instruction. Over a period of time, the impacts of testing are gradually taken for granted as parts of ordinary instruction. For example, teachers would not recognize that solving arithmetic problems arranged horizontally is an effect of external testing. Test publishers save space and money by presenting problems in this way. Alert curriculum developers found that pupil performance reliably differed according to whether the same addition problem was presented horizontally or vertically, and, to give pupils practice for the test, began including horizontal problems in texts and worksheets. Another example in math instruction is the timed test in problem-solving wherein pupils weekly take tests in solving as many arithmetic problems as they can within a one- or two-minute time limit. Yet another example of hidden effects of testing in math is the instruction pupils receive, grade after grade, in how to respond to story problems of a specific type. Teachers teach a set of steps in "problem-solving" that involve deciding from verbal clues, such as the words "all together," the correct algorithm to use, then do the arithmetic and supply the correct label. Romberg and Zarinnia (1989) pointed out that this definition of problem-solving was impoverished and rarely representative of authentic problems in mathematics.

Other examples of testing effects hidden in ordinary instruction are Systematic Review, seat work in the Reading Mastery curriculum in which pupils practice test-like exercises, and teaching logical operations as a way of boosting comprehension scores. Textbook publishers select content and problem formats in part by looking at test items, for example items that ask what numbers make a number sentence true. The inclusion of Study Skills as a part of curriculum that the district requires is another example of hidden testing effects. Teachers accept it as a legitimate part of the school curriculum and make room for teaching study skills in a crowded program. Yet at least two-thirds of the material in the Handbook relates to teaching test-taking skills or drilling on maps and graphs that the ITBS covers. Indeed, the reading of graphs and maps recurs in grade after grade and repeats in math, social studies, science, and study skills. New teachers have no idea that such material results from anything but reasoned debate about what schools naturally ought to teach. The similarity in content and format between the ITBS, BST, and CUES represents another hidden effect of testing in ordinary instruction. The alignment of Scope and Sequence and CUES in language arts to the ITBS represents another structural effect of external testing on curriculum.

By teachers' definitions, external testing affects pupils. For pupils, particularly younger ones, most teachers believe that standardized testing is "cruel and unusual punishment." Because of the length and difficulty of tests, the number of tests, the time limits, the restrictions and "individualistic" nature of test-taking, the fine print, and difficulty in transferring answers to answer sheets, teachers believe tests cause stress, frustration, burnout, fatigue, physical illness, misbehavior and fighting, and psychological distress. Some teachers believe that the tests cause subsequent test anxiety and failure mentality. Teachers believe many pupils simply guess or give up trying to perform when they encounter items that are too difficult for them and worry that test scores will determine their course grades or promotion.

Although other interest groups fail to confirm the beliefs of teachers in this study, teachers that Nolen surveyed (Nolen et al., 1989) shared the beliefs of our teachers. Lacking the kind of evidence needed, which is inaccessible to observers, we are unable to resolve this issue.

We contend that most of these effects occur not because of the norm-referenced characteristics of the ITBS, but because of the power of the test to evoke consequences (the pressures teachers transfer from administrators to pupils) and because of the number of tests pupils must take.

External testing affects teachers. Teachers' view of the deficiencies of achievement tests notwithstanding, they feel ashamed and embarrassed if their pupils score low or fail to "grow" by district standards. They feel relieved rather than proud when scores are high, for they know that test scores are weighted more by pupils' socioeconomic status and level of effort than anything teachers personally do in the classroom. The chagrin they feel comes from their well-justified belief that audiences external to the school lack interpretive context and attribute low scores to lazy teachers and weak programs.

Data we pieced together from multiple sources suggest that external testing also diminishes teachers' sense of efficacy and perhaps, over time, their competence as well. First, the three achievement testing standards the district adopted (ITBS grade equivalent/grade placement standard, the ITBS growth metric, and the BST 75 percent mastery standard) correlate poorly with one another. Even after following the district Scope and Sequence, teachers find it nearly impossible to satisfy all three standards without blatant teaching to the tests (and maybe not even then). Thus, teachers will likely look bad and need to defend themselves based on one or more of these criteria, depending on whether district administrators or board members choose to single them out. Second, external audiences tend to focus attention on random differences and trends, which because they are unstable and unreal, are outside teachers' control. Third, the tests themselves are less reliable and the variances higher than most people recognize, and they correlate substantially with several pupil characteristics other than program or teacher quality. Thus, teachers are always kept off-balance and feeling inept in the face of the "magic numbers." To make the numbers come out just right is to ask the impossible of teachers and principals. Successful teachers are those that seek alternative indicators of effectiveness and recognition.

As teachers take more time for test preparation and align instruction more closely with test content and form, they diminish the range of instructional goals and activities they know about or practice. They forget that problem-solving may be a broader concept than the operations necessary to solve word problems. They forget that reading is extracting meaning from text rather than correct performances of subskills of reading. They learn less science and social studies, because the authority is the text and the test, and nothing further in these subjects is expected of them. Since tests do not measure disciplined inquiry, integration of knowledge, production of discourse on novel problems (Romberg & Zarinnia, 1989), critical thinking, civic participation, and cultural knowledge, teachers ignore these attainments and later lose the capacity to produce or even imagine them.

One criterion on which teachers are evaluated is the extent to which they are "teaching the adopted curriculum," and the basis of teacher supervision is one of conformity to centrally defined standards of teaching behavior. As we have shown, the curriculum is packed and geared to external tests, leaving little room for innovation, divergence, adaptation to local circumstances and needs, and teacher choice. Teachers' sense of themselves as autonomous professionals and authorities on curriculum and instruction is constrained. When all instructional decisions are controlled from the district office, teachers may lose the capacity to define attainment for themselves or accomplish it in their classrooms. As choices of what to teach are made elsewhere and required methods grow increasingly test-like, teachers' work is deskilled and degraded. Teachers who take into account prescriptions that they can only read from scripts and manuals and correct pupils' worksheets are less likely to define themselves as competent to teach by other means. Overall capacity of our schools is likely to decline.

Overall Research Perspective

As with any study of the scope and complexity of this one, there are many ways to look at the resulting data and many possible interpretations that we or our readers might draw from them. In the course of this research, we have come to define the role of external testing in public schools as a problem in micro-politics, a part of symbolic interactionism that stresses conflict. Following Ball (1987), we see the school as an "arena of conflict" in which various interest groups dynamically compete for relative autonomy, material resources, and influence. Manipulating the symbols offered by external testing programs is one tactic used by interest groups. Although we are indebted to the various experts in psychometrics, particularly Mehrens and Kaminski, Shepard, Haladyna, and Berk, for some insights into the nature of external testing programs, we have come to realize that to define the problem of the role of testing as solely psychometric is to oversimplify. But it is the psychometric weaknesses of external testing programs (their poverty of representing broad concepts of educational attainment, their corruptibility, their crudeness and instability) that make them such handy weapons in skirmishes among interest groups.

Enforcing test score standards, prescribing the packed curriculum, and practicing a method of supervision that emphasizes compliance are three interlocking means by which district administrators attempt to increase their power and relative autonomy, thereby reducing the relative autonomy of teachers and principals. By these means central administrators assert a particular ideology or definition of schooling and urge others to accept it. That definition of schooling is one in which major decisions about what to teach and how to teach it are made centrally and filter down a hierarchy of authority. In the district we studied, the centralized curriculum is a hierarchically-arranged construction of basic skills, repetitiously drilled and repeated across grades, buttressed by a set of criterion-referenced mastery tests build around a common set of goals and content. The ITBS is the ultimate, though not the sole target. The rhetoric of the administration allows individual schools some autonomy over methods and approaches but not over the selection and sequence of goals and content, nor to define standards of attainment. Goals and standards, however, are implicit in methods, so that commitment of schools to alternative methods puts them at odds with what the district expects them to achieve.

Principals are active participants in the conflict, and in fact engage in complementary tactics, attempting to impose their own definitions of the school upon teachers and external audiences, using what means they have to increase their autonomy, and manipulating test score symbols and other symbols of their schools' attainment. Hamilton's principal negotiated a substitution of scores on the Metropolitan, massaged to account for transience rates and initial reading levels of pupils. Both principals deflected criticisms by pointing to the growth standard which their schools met, at least one of them privately acknowledging the psychometric weakness of the metric. Both sought awards and recognitions for their schools and allies in professional associations and the community. Test score standards are also tactics district administrators use in the management of relationships with external audiences. High external test scores protect the district's range of autonomous actions from intrusions by state and federal government and special interest groups.

It is our contention that no test score ever improves schools. Attempting to improve schools by boosting scores or to reform schools by shaming them with low rankings can only achieve short-term, largely symbolic changes that will not generalize to alternative indicators. When society's interest in education becomes focused on test scores, better schools will not result; rather, the schools will suffer a decreased capacity for conveying worthwhile curricula and reaching worthy goals. What schools do and what they produce will be obscured in a fog of misinformation.

References

American Psychological Association, American Educational Research Association, National Council on Measurement in Education. (1985). *Standards for educational and psychological testing.* Washington, DC: American Psychological Association.

Airasian, P., & Madaus, G.F. (1983). Linking testing and instruction: Policy issues. *Journal of Educational Measurement, 20,* 103-118.

Baker, E.L. (1989, March). *Cannell revisited: Accountability, test score gains, normative comparisons, and achievement.* Paper presented at the Annual Meeting of the American Educational Research Association, San Francisco.

Ball, S.J. (1987). *The micro-politics of the school.* London: Methuen.

Bangert-Drowns, R.L., Kulik, J.A., & Kulik, C.C. (1988). *Effects of frequent classroom testing.* Unpublished manuscript, University of Michigan, Ann Arbor.

Berk, R.A. (1988). Fifty reasons why student achievement gain does not mean teacher effectiveness. *Journal of Personnel Evaluation in Education, 1,* 345-363.

Bishop, C.D. (1988). *Statewide report for Arizona pupil achievement.* Phoenix: Arizona Department of Education.

Bracey, G.W. (1987). Measurement-driven instruction: Catchy phrase, dangerous practice. *Phi Delta Kappan, 68,* 683-686.

Briggs, P.W. (1987). *Phoenix—An urban city: A descriptive report on urban characteristics.* Tempe, AZ: College of Education, Arizona State University.

Campbell, D.T. (1979). Assessing the impact of planned social change. *Evaluation and Program Planning, 2,* 67-90.

Cannell, J.J. (1987). *Nationally normed elementary achievement testing in America's public schools: How all fifty states are above the national average.* Daniels, WV: Friends for Education.

Cohen, S.A. (1987). Instructional alignment: Searching for a magic bullet. *Educational Researcher, 16,* 16-19.

Cook, T.D., & Campbell, D.T. (1979). *Quasi-experimentation: Design & analysis for field settings.* Chicago: Rand McNally.

Corbett, H.D., & Wilson, B.L. (1989). *Raising the stakes in statewide mandatory minimum competency testing.* Philadelphia: Research for Better Schools.

Cronbach, L.J. (1984). *Essentials of psychological testing.* New York: Harper and Row.

Cronbach, L.J. (1971). Test validation. In R.L. Thorndike (Ed.), *Educational measurement* (2nd ed., pp. 443-507). Washington, DC: American Council on Education

Cronbach, L.J., & Furby, L. (1970). How should we measure "change"—or should we? *Psychological Bulletin, 74,* 64-80.

Darling-Hammond, L., & Wise, A.E. (1985). Beyond standardization: State standards and school improvement. *Elementary School Journal, 85,* 313-336.

Deaton, W.L., Halpin, G., & Alford, T. (1987). Coaching effects on California Achievement Test scores in elementary grades. *Journal of Educational Research, 80,* 149-155.

Dorr-Bremme, D.W., & Herman, J.L. (1986). *Assessing student achievement: A profile of classroom practices* (CSE Monograph Series in Evaluation No. 11). Los Angeles: UCLA Center for the Study of Evaluation.

Edelman, J. (1981). The impact of mandated test program on classroom practices: Teacher perspectives. *Education, 102,* 56-59.

Edelman, M. (1976). *The symbolic uses of politics.* Urbana, IL: University of Illinois Press.

Eisner, E. (1981). On the differences between scientific and artistic approaches to qualitative research. *Educational Researcher, 10,* 5-9.

Ellwein, M.C. (1987). *Standards of competence: A multi-site case study of school reform.* Unpublished doctoral dissertation, University of Colorado, Boulder.

Erickson, F.E. (1986). Qualitative methods in research on teaching. In M. Wittrock (Ed.), *Handbook of research on teaching* (3rd ed., pp. 119-161). New York: Macmillan

Feiman-Nemser, S., & Floden, R.E. (1986). The cultures of teaching. In M. Wittrock (Ed.), *Handbook of research on teaching* (3rd ed., pp. 505-526). New York: Macmillan.

Ginsberg, P.E. (1984). The dysfunctional side effects of quantitative indicator production: Illustrations from mental health care. *Evaluation and Program Planning, 7,* 1-12.

Glaser, B.G. (1978). *Theoretical sensitivity.* Mill Valley, CA: The Sociological Press.

Glaser, B.G., & Strauss, A.L. (1967). *The discovery of grounded theory.* Chicago: Aldine Press.

Glaser, R. (1984). Education and thinking: The role of knowledge. *American Psychologist, 39,* 93-104.

Glass, G.V. (1989). Using student test scores to evaluate teachers. In J. Millman & L. Darling-Hammond (Eds.), *The new handbook of teacher evaluation.* Newbury Park, CA: Sage.

Gold, R.L. (1958). Roles in sociological field observations. *Social Forces, 36,* 217-223.

Haas, N.S., Haladyna, T.M., & Nolen, S.B. (1989). *Standardized testing in Arizona: Interview and written comments from teachers and administrators* (Tech. Rep. No. 89-3). Phoenix: Arizona State University West Campus.

Haladyna, T.M., Haas, N.S., & Nolen, S.B. (1989). *Test score pollution* (Tech. Rep. No. 89-1). Phoenix: Arizona State University West Campus.

Iverson, G. (1984). Raising test scores. *Educational Measurement: Issues and Practice, 3,* 45-46.

Jordan, P.W. (1987). School faculty disciplined over tests: School teachers draw transfers. *Fairfax Journal, 48*(3), A-1, A-5.

Kellaghan, T., Madaus, G.F., & Airasian, P.W. (1982). *The effects of standardized testing.* Boston: Kluwer-Nijoff.

Kulik, J.A., Kulik, C.C., & Bangert, R.L. (1984). Effects of practice on aptitude and achievement test scores. *American Educational Research Journal, 21,* 435-447.

Linn, R.L., Graue, M.E., & Sanders, N.M. (1989, March). *Comparing state and district test results to national norms: Interpretations of scoring "above the national average."* Paper presented at the Annual Meeting of the American Educational Research Association, San Francisco.

Madaus, G. (1987). *Testing and the curriculum.* Chestnut Hill, MA: Boston College.

Mathison, S.M. (1987). *The perceived effects of standardized testing on teachers and curricula.* Unpublished doctoral dissertation, University of Illinois, Champaign.

Mayeski, G.W. (1973). *A study of the achievement of our nation's students.* Washington, DC: Office of Education, U.S. Department of Health, Education, and Welfare.

McLean, L. (1989, March). *Technical and social issues in subject matter theories that guide curriculum and testing.* Paper presented at the Annual Meeting of the American Educational Research Association, San Francisco.

McCracken, G. (1988). *The long interview.* Newbury Park, CA: Sage.

Mehrens, W.A., & Kaminski, J. (1988). *Using commercial test preparation materials for improving standardized test scores: Fruitful, fruitless, or fraudulent?* East Lansing, MI: Michigan State University School of Education.

Messick, S. (1988). Validity. In R.L. Linn (Ed.), *Educational measurement.* Washington, DC: American Council on Education.

Miles, M.B., & Huberman, A.M. (1984). *Qualitative data analysis: A sourcebook of new methods.* Beverly Hills, CA: Sage

Mishler, E.G. (1986). *Research interviewing.* Cambridge, MA: Harvard University Press.

Nolen, S.B., Haladyna, T.M., & Haas, N.S. (1989). *A survey of Arizona teachers and school administrators on the uses and effects of standardized achievement testing* (Tech. Rep. No. 89-2). Phoenix: Arizona State West Campus.

Pechman, E.M. (1985). Cheating on standardized tests: Why does it exist? In P. Wolmut and G. Iverson (Eds.), *National Association of Test Directors 1985 symposium.* Portland, OR: Multnomah ESD.

Peterson, P.L. (1989). Alternatives to student retention. In L.A. Shepard & M.L. Smith (Eds.), *Flunking grades: Research and policies on retention.* London: Falmer Press.

Phillips, D.C. (1989, March). *Validity in qualitative research.* Paper presented at the Annual Meeting of the American Educational Research Association, San Francisco.

Polkinghorne, D.E. (1988). *Narrative knowing and the human sciences.* Albany, NY: State University of New York Press.

Popham, W.J. (1987). The merits of measurement-driven instruction. *Phi Delta Kappan, 68,* 679-682.

Price, H.H. (1969). *Beliefs.* London: Allen and Unwin.

Rein, M. (1976). *Social science and public policy.* New York: Penguin Books.

Resnick, L.B., & Resnick, D.P. (1989). Assessing the thinking curriculum: New tools for educational reform. In B.R. Gifford & M.C. O'Connor (Eds.), *Future assessments: Changing views of aptitude, achievement, and instruction.* Boston: Kluwer Academic Publishers.

Romberg, T.A., & Zarinnia, E.A. (1989). *The influence of mandated testing on mathematics instruction: Grade 8 teachers' perceptions.* Madison, WI: National Center for Research in Mathematical Science Education, University of Wisconsin.

Rothman, R. (1989, October 11). Learning goals said to demand better assessment. *Education Week, 8,* 12.

Samson, E. (1985). Effective training in test-taking skills on achievement test performance: A quantitative synthesis. *Journal of Educational Research, 78,* 261-266.

Scruggs, T.E., White, K.R., & Bennion, K. (1986). Teaching test-taking skills to elementary-grade students: A meta-analysis. *The Elementary School Journal, 87,* 69-82.

Seidel, J., Kjolseth, R., & Clark, J. (1985). *The Ethnograph: A user's guide.* Littleton, CO: Qualis Research Associates.

Shanker, A. (1988, April 24). Exams fail the test. *The Sunday New York Times.*

Shepard, L.A. (1989, March). *Inflated test score gains: Is it old norms or teaching the test?* Paper presented at the Annual Meeting of the American Educational Research Association, San Francisco.

Smith, M.L. (1982). *How educators decide who is learning disabled.* Springfield, IL: Charles C. Thomas Press.

Smith, M.L., & Shepard, L.A. (1988). Kindergarten readiness and retention: A qualitative study of teachers' beliefs and actions. *American Educational Research Journal, 25,* 307-333.

Strauss, A.L. (1987). *Qualitative analysis for social sciences.* Cambridge, MA: Cambridge University Press.

Stryker, S. (1980). *Symbolic interactionism.* Menlo Park, CA: Benjamin/Cummings.

Sullivan, H.S. (1954). *The psychiatric interview.* New York: W.W. Norton.

Wilson, B.L., & Corbett, H.D. (1989). *Two state minimum competency testing programs and their effects on curriculum and instruction.* Philadelphia: Research for Better Schools.

Appendix A

Summary of a Survey of Arizona Educators

A trio of technical reports (Haas, Haladyna, & Nolen, 1989; Haladyna, Haas, & Nolen, 1989; Nolen, Haladyna, & Haas, 1989) presented the results of a project sponsored by the Arizona Department of Education. The motivation for commissioning the report was the proposed legislation to alter the state's program of mandated assessment. Prior to 1988, the legislative mandate was to test every pupil every year from first to eighth grades on the Iowa Test of Basic Skills, the Stanford Achievement Test in grade nine, and the Stanford Test of Academic Skills in grades 10-12. Legislation passed in 1988 abolished the mandate in grades one and twelve and left testing in those grades to the decision of school districts. Proposed legislation called for norm-referenced standardized achievement in benchmark years only and in samples of pupils. Criterion-referenced tests based on the Arizona Essential Skills would be developed at the state for all pupils. To assist legislators in their deliberation over the proposed changes, an evaluation of the current testing program was proposed. Professors in the Department of Education and Human Services at Arizona State University, West Campus received the contract to do the study. Permission to reference the technical reports was given by Dr. Haladyna.

The researchers surveyed teachers and administrators throughout the state using a two-stage sampling design. From a random start on a sampling frame of schools, every seventh school was chosen. The principals of the schools chosen were sent packets of questionnaires to pass out to all teachers employed there. Of the 5,770 questionnaires sent out, 41 percent responded. Forty-seven percent of administrators also returned questionnaires. It was not deemed possible to do follow-up studies of those who failed to respond. The demographic data of responders were analyzed and seemed to match demographic data known to characterize the population.

Items on the questionnaire fell into four categories: uses of standardized achievement testing, beliefs about standardized testing, test preparation activities, and perceived effects of testing.

In addition, interviews were conducted with a teacher and from one to three teachers in each of 30 school districts reflecting a cross-section of the state. Interviews of 15-20 minutes were conducted to corroborate the findings of the questionnaire study. Content analysis of interviews as well as the open-ended comments gathered on the questionnaires was performed.

Uses of test results. Among the many findings relevant to elementary staff were these: Only about one third of the teachers report using the ITBS to guide instruction, diagnose learning problems, communicate with parents, place students for instruction, or evaluate programs or curriculum. At least half report using the test scores to identify gifted or remedial students. About 40 percent believe administrators "routinely or often" use ITBS scores to evaluate teachers, curricula, and school effectiveness. Sixteen percent of the teachers say administrators use scores to determine tenure and merit pay. A majority of teachers also believe that their districts and school boards use scores to advertise the school and evaluate district effectiveness. A majority of responding teachers believe that the state uses ITBS in school competitions, to evaluate school, district, and state effectiveness, to create political pressure and lobby for or against funding for education. In general, administrators reported their own use of scores was less than what teachers believed it was. Teachers were also asked about which uses of test scores they endorsed.

What uses they believed were appropriate were less than actual uses, in almost all categories. In other words, although 42 percent of the teachers believed that administrators use scores to identify teachers' strengths and weaknesses, only 8 percent believed this was an appropriate use.

Beliefs about testing. In their beliefs about testing, the Arizona elementary teachers were uniformly pessimistic about what scores reveal. Only 16 percent felt that ITBS scores reflect a single year's learning. About one third felt that the scores reflect a cumulative attainment over the pupil's entire career. Only three percent felt that the tests were accurate for minorities or non-English speakers. When asked what factors affect ITBS scores, 70 percent named family background, 82 percent nominated student effort and family support for learning, 55 percent named class size (25 percent of the administrators). Only 40 percent of the teachers (compared to 68 percent of the administrators) named teacher skill as a factor in affecting ITBS scores. Thirteen percent agreed with the statement that the benefits of testing outweigh its drawbacks. When asked how frequently the state should require the administration of standardized tests, 18 percent chose the options of once or twice each year, while 63 percent chose the option of three to five times between second and eleventh grades. Except for the items mentioned above, administrators' beliefs about testing mirrored those of the teachers.

Test preparation. Eighty percent of the teachers said they were encouraged to raise test scores. Only seven percent reported that they are urged to prepare their pupils by teaching actual test items. Two thirds of the respondents are encouraged to focus on skills they know will be tested and use the same format on their classroom tests that they know the ITBS uses. Three-quarters of them report being encouraged, usually by principals or district administrators, to teach the techniques of test-taking. What the teachers actually report doing is demonstrating marking procedures (69 percent), give general tips on test-taking and discuss the importance of the test (70 percent), encourage attendance (93 percent), use commercial test-preparation packages (41 percent), teach or review topics covered by the test (66 percent), teach vocabulary that will be on the test (26 percent), teach actual items from last year's or current test (10 percent), and teaching techniques of taking tests (60 percent).

Twenty-eight percent of the teachers report that they started preparing for the tests two or more months prior to the test, about equally divided into frequencies of either daily, weekly, or less often. Twenty-two percent say they start the week before the test, most working daily.

During the test week, teachers report that it is common or very common to provide students with snacks (38 percent), do more test-taking practice (90 percent), review skill that will be covered on the next day's test (44 percent), and give rewards for completing the test (14 percent). Ninety-five percent of the respondents say they follow the test directions exactly. Eight percent admit increasing or decreasing the time allotted. Eighty-eight percent say that test security is adequate. Fifty percent say they spend either four or five days of the testing week on non-instructional activities.

Effects of testing on pupils. Asked to list the symptoms of students during the test, the following percentages of elementary teachers responded that "every year" or "usually" they say truancy (15 percent), stomach symptoms (29 percent), vomiting (8 percent), crying (21 percent), irritability (38 percent), wetting or soiling themselves (7 percent), too many rest room breaks (29 percent), excess concern over time limits (44 percent), freezing up (41 percent), headaches (40 percent), hiding (8 percent), refusing to take the test (10 percent), and increased aggression (33 percent). Except for truancy, administrators reported seeing these symptoms at lower rates.

In their conclusions and recommendations, the authors noted that the ITBS has not been validated for the purpose that Arizona uses it. They agreed with the respondents that its limited validity and utility make testing every pupil in every year a policy that costs more than it is worth. They accept the beliefs of the educators that pupils are deleteriously affected by taking the test. They recommend periodic, benchmark testing by the ITBS on a random sampling basis and support the development and mandated administration of criterion-referenced tests of Arizona Essential Skills. For the ITBS, the authors recommend that using commercial materials such as *Scoring High* be outlawed, for they constitute unethical practice and further "pollute" the inferences that can legitimately be drawn from test data.

For purposes of the present, qualitative study, the survey provides a means of placing the practices and beliefs of Hamilton and Jackson educators into a range of cases and beliefs. It confirms the interpretations and extends many of the inferences to known, arguably representative samples within Arizona.

Appendix B

These are the views of The Phoenix Gazette as an institution. Not signed by an individual writer.

Editorials

These are the views of The Phoenix Gazette as an institution, thus are not signed by an individual writer.

Test scores disappointing

State Superintendent of Public Instruction C. Diane Bishop used the release of 1989 statewide standardized test results to lobby for a new testing program.

While there is nothing wrong with adding other measures to the legislatively mandated program already in place, it should not be weakened further. Last year Bishop led the charge to exempt first and twelfth graders.

This year's disappointing results demonstrate the value of standardized testing. Students in grades 1-12 tested below the national average in mathematics in all but one grade, below average for reading in seven grades and below average for rammar in five grades.

Most disturbing, the worst showing ocred among students in the first three es, despite additional legislative appropriations. According to Bishop, this year's first-time use of new tests and higher national norms explain the declines.

However, the scores of seventh and eighth grade students held steady, indicating that the decline in the early grades probably is significant.

Although further analysis is necessary, since the new, higher 1988 norms reflect gains in national achievement since 1985, the scores appear to indicate that Arizona elementary students are not keeping pace with their peers nationally.

All parents want to be able to measure their children's progress against a national benchmark. An Arizona-only test will not provide that important information.